is little reason to think explicitly about requirements. However, any project benefits from the discipline of being more systematic about its purposes and their priorities. In addition, being explicit about requirements enables traceability and impact analysis. Traceability means being able to relate an interaction or feature of a system to the requirement it satisfies; impact analysis runs the causal link between requirements and features in the opposite direction to assess what or who will be affected if requirements change.[1]

Requirements for Interactions

When we describe interactions in a generic or broad way as we did in *Activities in Organizing Systems* we see that all organizing systems have some common interactions, but most of the time we want to pay attention to the more specific interactions that are designed to create value in a particular organizing system because of the kind of resources it contains (“Interaction and Value Creation”). The domain, scope, and scale of the organizing system determines which interactions are possible and which ones must be explicitly supported, but the priorities of different interactions are more often determined by decisions about intended users. (See “Number and Nature of Users”.)

For most organizing systems other than personal ones, the set of interactions that are implemented in an organizing system is

1. An easy to remember framework for prioritizing requirements is MoSCoW, which classifies them as Must, Should, Could, and Won’t (Desoky 2010). (Winkler and Pilgrim 2010) is a comprehensive review of academic research and best practices for requirements traceability.

strongly determined by business model considerations, funding levels, or other economic factors. For-profit firms often differentiate themselves by the number and quality of the interactions they support with their resources, some by supporting many of them and some by supporting a minimal number. This differentiation is strongly shaped by and also shapes user preferences; some people prefer self-service or unmediated interactions, while others prefer full service and mediated interactions.[2] Non-profit institutions like public libraries and museums are also subject to these constraints, but unfortunately they have fewer options for adjusting service levels or changing their targeted user populations when their funding is reduced.[3]

2. "Customer segments" or "customer models" are well-established constructs in product and service marketing and operations (Batt 2000) (Zeithami, Rust, and Lemon 2001). They are key parts of strategies for acquiring customers, increasing market share, and retaining customers. Customer segments can be identified using numerous overlapping criteria, including demographic variables, product or behavior choices, and preferred interaction locations or channels. For example, an airline might segment its customers according to their age, gender, home airport, ticketing class, and travel frequency.

3. Funding cuts for public libraries lead to reduced staffing, reduced hours, and reduced acquisitions and many of them serve populations facing economic challenges of their own. (Johnson 2010).

Some requirements for interactions come along with technology requirements, to have resources in a particular format, to conform to a particular specification or standard in order to operate in some technology environment, or to interoperate with other parties or their organizing systems.

An essential requirement in every organizing system is ensuring that the supported interactions can be discovered and invoked by their intended users. In organizing systems with physical resources, good designers enhance the inherent affordances of resources with navigation and orientation aids that direct users to points of interactions ("Affordance and Capability"). With digital resources and information-intensive organizing systems, interactions are not immediately perceivable, and poor design can create overly complicated user interfaces in which many interactions are never discovered and thus never used.

It is tricky to compare the overall capabilities of organizing systems in terms of the number or variety of their interactions because what matters more is how much value they create. Organizing systems with active resources can create value on their own without an explicit user interaction ("Active or Operant Resources"). Other organizing systems exploit stored, computed, or *contextual information* to create value by eliminating the need for user interactions, such as location-based smartphone apps that push information to you when you are near some particular location or some person you know ("Affordance and Capability").[4]

4. Organizing systems differ in the extent they can initiate interactions or use information to make them unnecessary. In libraries the organizing systems are typically designed not to preserve user activity records longer than absolutely necessary; in commercial

About the Nature and Extent of Resource Description

Interactions with resources within an organizing system often depend on descriptions of individual resources or descriptions of the collections that contain them. In the bibliographic domain, generic or common interactions make use of descriptions that can be associated with almost any type of resource, such as the name, creator, and creation date.[5]

For example, any resource with a sortable name or identifier can be arranged alphabetically to enable it to be easily found, and any resource with a creation date can be discovered by a “what’s new” query to a resource collection.

Different types of resources must have differentiating properties, otherwise there would be no reason to distinguish them as different

organizing systems, user activity records are the basis of business processes that create highly detailed user models (called “microsegments” or “microcategories”) that enable personalized product and service offerings See (Taylor and Raden 2007), (Rosen 2012).

5. The Dublin Core was proposed in 1995 as a small vocabulary with 15 common elements that could be broadly applied. The emergence of many specialized derivatives of the Dublin Core since then illustrates the inherent tension between the simplicity of using a small set of common descriptive elements and the precision enabled by a large or more domain specific vocabulary.

types. These resource properties can be recorded in the terms of a description language to support one or more interactions or to answer one or more questions. Simply put, choices about the nature and extent of resource description depend on which interactions or questions are most frequent or important (“Describing Instances or Describing Collections”). If a particular property of a resource has no interactions that depend on it, there is no need to describe it. However, if an interaction depends on a description of a particular resource property, a missing description or one of inadequate precision and granularity means that the interaction will be impossible or inefficient to carry out because the resource will need to be further analyzed to create or extract the required description. An ISBN is a sufficient description to find a book in a directory, but if the ISBN is the only description associated with the book you will not be able to tell who wrote it. The tradeoffs imposed by the extent and timing of resource description have been a recurring theme in this book, with the tradeoff between recall and precision being the most salient (“When Is It Being Organized?”, “Affordance and Capability”, “Category Abstraction and Granularity”, “The Recall / Precision Tradeoff”).

The properties of resources that are easiest to describe are not always the most useful ones, especially for information resources. Anyone can determine the number of pages in a book, but often only a skilled cataloger can accurately describe what the book is about, a far more important property. (““Description” as an Inclusive Term” and “The Limits of Property-Based Categorization”) For non-text information resources this problem is magnified because the content is often in a semantically opaque format that it optimized for the devices that creates and processes it but which cannot usefully be analyzed by people. (“The Semantic Gap” and “Describing Non-text Resources”)

Business strategy and economics strongly influence the extent of resource description. In many museums and archives there are not enough trained people and time to describe every pottery fragment or document, and many resources are described only at an aggregate level. In contrast, some people argue that the explosion of content in physical and digital form mandates significant investment in descriptions that facilitate resource discovery in a crowded marketplace.[6]

Automated and computerized processes can create the resource descriptions in an organizing system and their use is primarily driven by scale ("Automated and Computational Resource Description"). Search engines index web pages and analyze their link structures because it would be impossible to treat the web as a traditional library and organize it by human effort. The benefits of digital cameras, video recorders, and similar devices would be far fewer if people had to manually identify and describe each resource when creating it. Instead, these devices can automatically assign some contextual metadata. Similarly, competitive pressures on vendors to provide real-time and context-sensitive information services mandate automated collection of contextual information like location from mobile phones, portable book readers and tablet computers.

Color Coded Library

6. (Register and McIlroy 2012) http://themetadatahandbook.com/.

Because he is presumably familiar with the contents of all of his books, interaction designer Juhan Sonin organized his library according to their spine colors. This organizing principle is a highly individual and aesthetic one, but it would probably not appeal to people unfamiliar with the collection and would bring chaos to a library of larger size.

(Photo by See-ming Lee. Creative Commons CC-BY-SA-2.0 license.)

We might seek some optimal degree of description given some set of requirements or purposes for an organizing system and some estimate of the organizing effort that could be applied; in practice this is elusive for two reasons, both relating to scope and scale. First, as the number of users of an organizing system increases, it becomes more difficult to identify and anticipate all its possible purposes and constraints it must satisfy. Even if most users share the goals for the organizing system, any particular user might have some additional specialized use for some attributes or relationships that would require more description to satisfy.[7]

Second, even if it were possible to implement some optimal degree of description in a particular organizing system, we would still encounter problems when multiple organizing systems exist in the same domain or in domains that intersect across context, device, or application boundaries. Since organizing systems are designed and evolve to satisfy the specific requirements of their particular context, companies will often describe the same resources differently, which creates integration and interoperability problems when companies need to exchange and combine information resources ("Transforming Resources for Interactions").

About Intentional Arrangement

Organizing principles depend on resource descriptions, so requirements for the former are always intertwined with those for the latter. Specifying requirements for the intentional arrangement of resources is analogous to specifying why and how resource categories can be created ("Principles for Creating Categories"). In

7. For example, we have often used a home kitchen as a setting for organizing systems. Suppose the home kitchen is to be used as the set for a cooking show and the designers want to arrange cookbooks to make the background visually pleasing. The designers would like to search for cookbooks on the basis of size and spine color, but these descriptive elements would be of little value to other users.

turn, the creation of resource categories often becomes a question about the number and kind of resource properties that might be analyzed and exploited by organizing principles.

We noted that there is a continuum of category formation that ranges from minimal use of resource properties to more rigorous use of multiple properties, and finally to statistical or composite use of multiple properties, some of which are induced or inferred rather than explicit. The simplest principle for defining a category is by enumeration, just putting the resources into a set without any specification of any properties they might share. The enumerated resources might very well have common properties, but the principle of enumeration ignores them; the only property that matters for that principle is that the resources are in the same set. This corresponds to the simplest principle of intentional arrangement, that of *collocation*, just putting the resources in the same location without any additional organization.

Collocated resources often acquire some additional arrangement as a result of their use; consider how the books, papers, or other resources gathered for some writing project often end up in piles in your office or on your desk close to your work area. For a small collection, the proximity-to-use organizing principle is the easiest way to satisfy a requirement to minimize the time to find frequently used resources.

As we have often seen, the scope and scale of the organizing system is a dominant design consideration and it applies to principles of resource arrangement too. The *collocation* principle of arrangement is sufficient for small resource collections because it is not necessary to define the optimal organization if the time to find any particular resource is short even for an inefficient search method of scanning the entire collection. Using the extrinsic property of frequency of use makes search slightly more efficient, but only in organizing systems where the user population is small or interacts with the resources in similar ways. Otherwise, arranging resources to facilitate their frequent access for some users would hinder other users who never use them. Imagine if you shared your office desk with someone who works all night on other writing projects and leaves his frequently used resources in piles close to his work area—which becomes your work area in the morning.

Larger resource collections usually require multiple organizing principles to manage the complexity that emerges when more users and more varied interactions must be supported. A valet parking lot might organize cars only by size to make optimal use of limited space when parking and fetching them, but when cars are organized for sale they would be organized by price, performance, seating capacity, manufacturer, and many other properties. It is essential to establish the priority of users and interactions because these requirements determine the order in which the principles are applied to arrange the resources. This ordering creates a logical resource hierarchy that affects the efficiency of interactions and the maintenance of the organizing system over time.

Information resources are invariably challenging to arrange because their aboutness is not an easily perceived property and because of the open-ended purposes they can serve. Information collections with broad scope most often use a standard system of bibliographic classification ("Bibliographic Classification"). In contrast, special libraries have narrower collections that need to support domain-specific interactions for a relatively small set of users, and as a result they require more specialized organizational schemes.[8] The

8. The category of special libraries includes law libraries, corporate libraries—both those that support the head office and the research organization—medical libraries, military libraries, museum libraries, prison libraries, and

principles for resource arrangement in large firms of every type are often required to conform with laws and regulations for accounting, taxes, human resources, data retention and non retention, access control, and other functions. (See “Governance in Business Organizing Systems”, “Mandated Classifications”)

Dealing with Conflicting Requirements

Any individual, group, or enterprise can create an organizing system that meets their specific requirements, but once this organizing system involves two or more parties with different requirements, there is a potential for conflict. Roommates or spouses sometimes argue about how to organize items in the kitchen, in the refrigerator, or in some other shared space. To a person who arranges spices alphabetically and condiment jars by size, arranging them according to cuisine or frequency of use makes no sense. Similarly, if you are the sole user of a Dropbox or other cloud storage account, you can organize it any way you want. You can use any number of folders that need only make sense to you, or you can leave everything unorganized in a single folder. However, if you share the Dropbox account with another person, they are likely to have different organizational needs or preferences. Perhaps you tend to organize resources by file type, while they prefer to organize resources by topic or project.

A small number of people can often agree on an organizing system that meets the needs of each participant through informal negotiations. The potential for conflict increases when more people

might even be stretched to include libraries of software components.

are involved, and “bottom-up” *ad hoc* negotiations to resolve every disagreement between every pair of participants just are not feasible. In many domains conflicts are avoided or suppressed because the parties have developed or agreed to conform to standards (“Classification and Standardization”). Nevertheless, conflicts in organizing principles for large-scale organizing systems are often resolved by parties with the legal authority or economic power to impose a solution on all the participants in a “top-down” manner.[9]

9. Some people call this the “Wal-Mart approach” to standardization. A firm with dominant market power does not need to negotiate standards because it can impose whatever standards it chooses on its partners as a condition of doing business with them. When there are conflicting requirements, different relationships within the set of participants trying to reach agreement, and different extents to which they are subject to the authority behind the desired agreement, it is not surprising that approaches “that require perfect coordination and altruism are of no practical interest” (Rosenthal et al. 2004, p 47).

73. Designing and Implementing an Organizing System

Requirements define what must be done but NOT how to do it; that's the role of the design and implementation phases. Being explicit about requirements and the intended scope and scale of an organizing system before moving onto these phases in an organizing system's lifecycle avoids two problems. The first is taking a narrow and short-term focus on the initial resources in a collection, which might not be representative of the collection when it reaches its planned scope and scale. This can result in overly customized and inflexible resource descriptions or arrangements that cannot easily accommodate the future growth of the collection. A second problem, often a corollary of the first, is not separating design principles from their implementation in some specific environment or technology.

Choosing Scope- and Scale-Appropriate Technology

A simple organizing system to satisfy personal record keeping or some short-lived information management requirements can be implemented using folders and files on a personal computer or by using "off the shelf" generic software such as web forms, spreadsheets, databases, and wikis. Other simple organizing systems run as applications on smart phones. Some small amount of configuration, scripting, structuring or programming might be involved, but in many cases this work can be done in an *ad hoc* manner. The low initial cost to get started with these kinds of applications must be weighed against the possible cost of having to redo a lot of the work later because the resources and the resource descriptions might not be easily exported to new ones.

More capable organizing systems that enable the persistent storage and efficient retrieval of large amounts of structured information resources generally require additional design and implementation efforts. Flat word processing files and spreadsheets are not adequate. Instead, XML document models and database schemas often must be developed to ensure more control of and validation of the information content and its descriptions. Software for version and configuration management, security and access control, query and transformation, and for other functions and services must also be developed to implement the organizing system.

Technology for organizing systems will always evolve to enable new capabilities. For example, cloud computing and storage are radically changing the scale of organizing systems and the accessibility of the information they contain. It might be possible to implement these capabilities and services to an organizing system in an incremental fashion with informal design and implementation methods. If information models, processing logic, business rules and other constraints are encoded in the software without explicit traceability to requirements and design decisions the organizing system will be difficult to maintain if the context, scope or requirements change. This is why we have repeatedly emphasized the importance of architectural thinking about organizing systems, beginning in "The Concept of "Organizing Principle"" where we proposed that organizing principles should ideally be expressed in a way that did not assume how they would be implemented. (See also ""Information Architecture" and Organizing Systems", "Classification vs. Physical Arrangement", and "Introduction")

Architectural Thinking

Much of the advice about designing and implementing an organizing system can be summarized as "architectural thinking," introduced in "The Concept of "Organizing Principle"". The overall purpose of architectural thinking is to separate design issues from implementation ones to make a system more robust and flexible. Architectural thinking leads to more modularity and abstraction in design, making it easier to change an implementation to satisfy new requirements or to take advantage of new technologies or procedures. It is also important to think architecturally about the design of the vocabularies and schemas for resource description and of classification systems to leave room for expansion to accommodate new resource types ("Implementing Categories" and "Principles for Maintaining the Classification over Time"). Doing so is easier if the descriptions are logically and physically distinct from the resources they describe. A checklist the brings together useful principles and processes for architectural thinking from all parts of this book is in the nearby sidebar.

Principles and Processes
for Architectural Thinking

- Explicitly define the purposes and users of the organizing system, recognizing that users might not agree on purposes or their priorities ("Why Is It Being Organized?").

- Select resources ("Selection Criteria") and design interactions ("Interaction and Value

Creation”) to support the primary or highest priority users.

- Specify information and interaction requirements in a conceptual and technology-neutral way (“The Concept of “Organizing Principle””) that conforms as much as possible to domain standards, schemas, or vocabularies (“Classification and Standardization”).
- Implement user interactions with design patterns to make them more discoverable, usable, and effective (““Information Architecture” and Organizing Systems”).
- Follow principles for good names (“Choosing Good Names and Identifiers”) and good resource descriptions (“Principles of Good Description”).
- Make an informed decision about the tradeoff between flexibility and complexity; a simpler system might be easier to adapt (“Principles for Maintaining the Classification over Time”).
- Make design and technology decisions consistent with the expected life of the organizing system (“Operating and Maintaining an Organizing System”).

Stop and Think: What is a Library?

The word “library” has several meanings that differ in how much architectural thinking they embody. When you tell someone you will meet them at the library for a cup of coffee and a study session it is a specific physical place. At other times you might use a more abstract notion of a library as an organizing system with a predictable type of collection, resource arrangement, and supported interactions. Both meanings are important for a city creating a new library. How would you ensure that both are considered in an effective way?

Nevertheless, architectural thinking requires more careful analysis of resources and implementation alternatives, and most people do not think this way, especially for personal and informal organizing systems. You can imagine that someone might arrange a collection of paperback books in a small bookcase whose shelf height and width were perfectly suited for the paperbacks they currently own. However, this organizing system would not work at all for large format books, and a paperback could not be added to the collection unless one was purged from the collection. It would be more sensible to start with a bigger bookcase with adjustable shelves so that the organizing system would have a longer lifetime.

You might think that large institutional organizing systems would avoid these problems caused by tying a collection too tightly to the physical environment in which it is initially organized, but sometimes they do not. A famous example involves the art collection of the Barnes Foundation, which had to keep its paintings in the exact same crowded arrangements when the museum made a controversial move from a small building to a larger one because the donor had mandated that the paintings never be moved from their original settings. (See the sidebar, The Barnes Collection).

For digital resources, inexpensive storage and high bandwidth have largely eliminated capacity as a constraint for organizing systems, with an exception for *big data*, which is defined as a collection of data that is too big to be managed by typical database software and hardware architectures.[1] Even so, big data collections are often large but homogeneous, so their scale is not their most important challenge from an organizing system perspective (“Scope and Scale of the Collection”).

> The Barnes Collection
>
> Albert Barnes was a chemist who made a fortune inventing a preventive treatment for gonorrhea and who then amassed perhaps the greatest private art collection ever, one that contained over 800 paintings by artists like Picasso, Renoir, Matisse, van Gogh, and Cezanne. In 1922 Barnes built a museum for his collection in his residential neighborhood in Merion, PA, a suburb of Philadelphia. Barnes did not open his collection to the

1. Note that this definition does not include any specific size threshold, such as some number of terabytes (thousands of gigabytes). This allows the threshold size that makes a collection a big data one to increase as storage technology advances. It also recognizes that different industries or domains have different thresholds (Manyika et al 2011).

public and in his will mandated that the collection never be moved, loaned, or sold.

In the decades after Barnes died in 1951 the Merion museum needed extensive repairs and security upgrades, and some people suggested that its remote location and access restrictions jeopardized its financial viability. However, a proposal to relocate the collection to Philadelphia seemingly violated the terms of the Barnes will.

A legal fight dragged on for decades. Finally in 2004 a judge ruled that the collection could be moved to Philadelphia, but only if the new museum contained exact copies of the gallery rooms of the original museum and arranged the paintings exactly as they were in Merion. The new museum building, opened in 2012, is ten times larger than the old one, but the collection takes up the same space as it did in Merion. The other 90% of the building is occupied by an auditorium, offices, classrooms, a gift shop, and other space that contains none of the collection.[2]

2. The history of the collection and the legal battle are described in (Anderson 2003). A documentary film with a conspiracy perspective is (Argott 2009). See also BarnesFoundation.org.

Distinguishing Access from Control

Because large resource collections are often used for multiple purposes by many different people or projects, they illustrate another important architectural issue for collections of digital resources. A requirement for access to resources does not imply a need to directly own or control them, and information-intensive and web-based businesses have increasingly adopted organizing system designs that involve storage of digital resources in the cloud, licensing of globally distributed resources, and outsourcing of information services. Designs that use these architectural concepts can realize functional and quality improvements because the location and identity of the service provider is hidden by an abstraction layer (“Value Creation with Physical Resources”, “Distinguish Identifying and Resolving”). However, separating access from ownership has been a cultural challenge for some libraries and museums whose institutional identities emphasize the resources they directly control and the physical buildings in which they control them.[3]

Standardization and Legacy Considerations

As we noted with the Barnes Collection, a building becomes old and outdated over time. The technology used in digital organizing systems becomes obsolete faster than physical buildings do. The best way to slow the inevitable transformation of today’s cutting edge technology to tomorrow’s legacy technology is to design with standard data formats, description vocabularies and schemas, and

3. See (Sandler 2006), (Freeman et al. 2005).

classification systems unless you have specific requirements that preclude these choices.

Even a requirement to interoperate with an organizing system that uses proprietary or non-standard specifications can usually be satisfied by transforming from a standard format (“Institutional Semantics”, “Implementing Interactions”). Similarly, it is better to design the APIs and data feeds of an organizing system in a generic or standard way that abstracts from their hidden implementation. This design principle makes it easier for external users to understand the supported interactions, and also prevents disclosure of any aspects of resource description or organization that provide competitive advantage. For example, the way in which a business classifies products, suppliers, customers, or employees can be competitively important.

Two important design questions that arise with data transformation or conversion, whether it is required by a technology upgrade or an interoperability requirement, are when to do it and where to do it. The job of converting all the resources in a collection can typically be outsourced to a firm that specializes in format conversion or resource description, and a batch or pre-emptive conversion of an entire collection enables an upgraded or new organizing system to operate more efficiently when it is not distracted by ongoing conversion activity. On the other hand, if resources vary greatly in their frequency of use, a “do-it-yourself on-demand” method is probably more cost effective as long as the conversion does not impact the interactions that need to be supported.

74. Operating and Maintaining an Organizing System

After the organizing system has been designed and implemented it can be put into its operation and maintenance phases. We will look at these from two perspectives, first from the point of view of individual resources, and then from the point of view of the organizing system's design and implementation. These two perspectives are not always clearly distinguished. Curation, for example, is often used to describe actions taken to maintain individual resources as well as those that result in new arrangements of them.

Resource Perspective

Sometimes an organizing system is implemented with its organizing structures and relationships waiting to be populated by resources as they are acquired and described. The scope and scale of the organizing system shapes how the descriptions are created and how the descriptions are then used to assign resources to the logical or physical containers of the organizing system. The most important decisions to be made at this point involve determining an appropriate mix of methods for creating the resource descriptions, because their cost, quality, consistency, completeness, and semantic richness depends on which human or computational agents do the work ("Creating Resource Descriptions").

For web-based and consumer-focused organizing systems, it is tempting to rely on users to assign descriptions, tags, or ratings to resources ("Tagging of Web-based Resources"). Some of these systems attempt to improve the quality and precision of these descriptions by providing forms, controlled vocabularies, or suggestions. Finding a balance is tricky; too much direction and control is demotivating to uncompensated volunteer describers, and too little of it results in the proverbial "tag soup."

An essential operational and maintenance activity is evaluation of resource descriptions, first with respect to the time and process by which they are created, and second with respect to how and when they support the designed interactions ("Evaluating Resource Descriptions").

Some organizing systems are initiated with a fixed set of resources that will not change in any way. For example, in an archive as most narrowly defined, neither the individual resources nor the organization of the collection as a whole will change. If an archive of Abraham Lincoln letters is established, we know that Lincoln will never revise any letters or write any new ones, and any new classifications or descriptions devised by people studying the archive will not be used to rearrange the letters.

Most organizing systems, however, need to support ongoing interactions with a collection that changes over time as new resources enter the collection and old ones leave. These selection and collection management processes are explicit in libraries, museums and similar institutions that maintain collections to satisfy the changing needs and preferences of their user communities ("Looking "Upstream" and "Downstream" to Select Resources").

Properties, Principles and Technology Perspective

It is useful to consider how an organizing system as a whole is operated and maintained over time. We can analyze how the system's organizing properties, principles and technology might change, and we can roughly order different types of change according to their overall impact.

The most predictable maintenance activities for an organizing system with an expected long lifetime ("Expected Lifetime") are incremental changes in description vocabularies and classification schemes ("Principles for Maintaining the Classification over Time"). These need to evolve when new instances or contexts require additional properties to maintain the distinctions between types of resources, but the basic principles embodied in the organizing systems are not affected.

Incremental category maintenance takes place even in personal organizing systems where the categories are not always explicit. The collection of clothes in a college student's closet and the categories and properties for arranging them will change somewhat when he graduates and takes a job in a downtown office building where he needs to dress more formally than he did as a student. He will learn that despite the common term in the category name, "student casual" and "business casual" do not contain the same sets of resources.

Category maintenance is an ongoing activity in institutional organizing systems. The most commonly used bibliographic classification systems all have numbering and naming schemes that allow for subdivision and extension to create new subcategories to accommodate resources about new fields of knowledge and technology.

As another example, the *Association for Computing Machinery*(ACM) professional society created a keyword classification in 1964 to organize articles in its many publications, but relentless change in the computing field driven by Moore's Law has required the ACM to significantly revise the system almost every decade.[1]

In contrast, changes in business organizing systems are more likely to be driven by economic factors. Resource properties for managing collections of resource and the information that describes them often change over time as a result of new products and services, mergers and acquisitions, or refined customer segmentation. More substantial changes in business organizing systems reflect the need to comply with laws and regulations that impose new requirements for tracing money flows or transactions. These mandated classifications and processes might require new organizing principles, not just incremental properties ("Mandated Classifications").

The choice of implementation technology influences how easy it is to handle these types of changes in organizing systems. In databases this problem is known as "schema migration." With XML implementations, schemas can be designed with "extension" or "codelist" elements to enable changes that will not invalidate existing information. Business processes that are driven by "executable specifications" like the *Business Process Execution Language*(BPEL) can be easily modified because the BPEL XML instance is used to configure the software that carries out the process it describes.[2]

1. No classification scheme ever devised is as unstable as the ACM's because new computing concepts, technologies, and application areas are constantly emerging. Even the society's name seems outdated.

2. For a formal computer science treatment of BPEL see (Fu, Bultan, and Su 2004); for a commercial perspective see

Another very predictable type of activity over time with organizing systems is a technology upgrade that improves its quality or capabilities without affecting the organizing principles. A student might replace his handwritten lecture notes with typed notes on a laptop or tablet computer but not significantly change the way the notes are organized.

Institutional organizing systems are adopting tiered storage systems that automatically move resources between different types of storage media to meet performance, availability and recovery requirements. For example, firms with high financial impact of downtime like banks run critical organizing systems with copies in "failsafe" or "hot" modes that are synchronized with the production environments to prevent any interruptions in information access if the latter are disrupted. On the other hand, resources needed for regulatory compliance can be kept on lower cost disk storage.

The most challenging kinds of maintenance activities for organizing systems involve changes to the principles for arranging resources along with changes in the implementing technology. An example is the ambitious effort to introduce semantic web and linked data concepts in bibliographic organizing systems ("Bibliographic Organizing Systems", "The Semantic Web World"). And change comes faster to businesses than to libraries and museums. New technologies can have a disruptive impact on business organizing system, forcing major changes to enable strategy changes that involve faster finding, retrieval, or delivery of informational or physical goods.[3]

http://www.oracle.com/technetwork/middleware/bpel/overview/index.html.

3. This means that the organizing systems used by business

Sometimes major changes to organizing principles and technologies can be introduced incrementally, with changes “rolled out” to different sets of resources or user groups during a transition period. However, sometimes the changes are inherently ”all or none” because it is impossible to have two conflicting organizing systems operating in the same context. An easy to understand example of an organizing system that changed radically is the system governing which side of the road you drive on, which was changed in Samoa in 2009. (See the sidebar, Driving in Samoa).

Driving in Samoa

Whether you travel by bus, car, or bicycle, you always keep to one side of the road. The convention of driving on either the right side or the left side is a legal standard that everyone takes for granted. However, you must follow it to ensure safe driving and avoid collisions and crashes.

This standard of which side of the road you drive on was not decided arbitrarily, but rather, it was adopted as a result of history, convention, and the need for organization. If you were the only person to use the road, you could choose to travel on any side you wanted, even travel right down the middle. As soon as more than

applications more often employ configuration management, version control, model-based code generation, and other computing techniques that robustly support the need for qualitative changes in the organizing systems.

one person needs to use the same road, the risk of collisions compels the creation of a coordinating standard.

In 2009, the government of Samoa took the rare step of changing the side of the road standard from driving on the right to driving on the left. The original standard reflected the influence of German colonization in the early 1900s. However, Samoa is both geographically close to and economically intertwined with Australia and New Zealand, former British colonies that follow the British convention of driving on the left side. This proximity gives Samoa access to a nearby source of used cars that would be attractive to Samoa's relatively poor population. So, the Samoan government decided to use its authority to change the driving standard so that more of its people could afford to buy cars.

As one could imagine, this decision was not implemented without controversy and opposition. While the decision benefited people currently without cars, it negatively affected those who already owned them. After a switch like this, what happens to the current market value of the thousands of cars designed to drive on the right? Opponents also claimed that the switch would cause unprecedented safety hazards. If even a small fraction of drivers were not able to immediately get the hang of driving on the other side, the accident rate could increase tremendously. Imagine the current pool of buses designed with doors that open on the right hand side—would they now let passengers out in the middle of the street? Who would pay to have the buses modified to put doors on the left hand side?[4]

4. (Barta 2009).

75. Key Points in Chapter Eleven

- What three initial characteristics of an organizing system influence most of the decisions about that organizing system?

  Most of the specific decisions that must be made for an organizing system are shaped by the initial decisions about its domain, scope, and scale.

  (See “The Organizing System Lifecycle”)

- What is the effect of broad scope in an organizing system?

  The impact of broad scope arises more from the heterogeneity of the resources and users in a collection rather than from their absolute number.

  (See “Scope and Scale of the Collection”)

- What is a practical effect of increasing collection size?

  Larger collections need more people to organize and maintain them, creating communication and coordination problems that grow much faster than the collection.

  (See “Scope and Scale of the Collection”)

- How do you avoid problems of scope and scale?

  Standardization is the best way to prevent problems of scope and scale.

  (See “Scope and Scale of the Collection”)

- What is an effect of a heterogeneous user community?

Organizing systems in the same domain and with nominally the same scope can differ substantially in the resources they contain and the interactions they support if they have different categories of users.

(See “Number and Nature of Users”)

- How can designers mitigate possible negative consequences of the system and technology they design?

  Designers who recognize that their systems have real consequences for people should commit to measures of transparency and an ongoing process of negotiation that enables those affected to voice concerns related to any detrimental effects the technology might have on them and their communities.

  (See “Number and Nature of Users”)

- What is the single biggest factor affecting the implementation of interactions?

  For most organizing systems other than personal ones, the set of interactions that are implemented in an organizing system is strongly determined by economic factors.

  (See “Requirements for Interactions”)

- What is the most essential requirement of interactions?

  An essential requirement in every organizing system is ensuring that the supported interactions can be discovered and invoked by their intended users.

  (See “Requirements for Interactions”)

- What is the role of computational agents in the creation and consumption of resource descriptions?

Automated and computerized processes can create the resource descriptions in an organizing system and their use is primarily driven by scale.

(See "About the Nature and Extent of Resource Description")

- What is the relationship between organizing principles and resource descriptions?

  Organizing principles depend on resource descriptions, so requirements for the former are always intertwined with those for the latter.

  (See "About Intentional Arrangement")

- What characteristics of resource descriptions impede growth?

  Overly customized and inflexible resource descriptions or arrangements cannot easily accommodate the future growth of the collection.

  (See "Designing and Implementing an Organizing System")

- What are the advantages of architectural thinking?

  Architectural thinking leads to more modularity and abstraction in design, making it easier to change an implementation to satisfy new requirements or to take advantage of new technologies or procedures.

  (See "Architectural Thinking")

- What is big data?

  For digital resources, inexpensive storage and high bandwidth have largely eliminated capacity as a constraint for organizing systems, with an exception for *big data*, which is defined as a collection of data that is too big to be managed by typical database software and hardware architectures.

(See “Architectural Thinking”)

- What are the most predictable maintenance activities?

  The most predictable maintenance activities for an organizing system with an expected long lifetime are incremental changes in description vocabularies and classification schemes.

  Another very predictable type of activity over time with organizing systems is a technology upgrade that improves its quality or capabilities without affecting the organizing principles.

  (See “Properties, Principles and Technology Perspective”)

- What changes in organizing systems make maintenance especially difficult?

  The most challenging kinds of maintenance activities for organizing systems involve changes to the principles for arranging resources along with changes in the implementing technology.

  (See “Properties, Principles and Technology Perspective”)

- What six questions should be asked when approaching any organizing system?

  What resources are being organized? Why are the resources being organized? Who does the organizing? When are the resources organized? Where are the resources organized? How much are the resources organized?

  (See the case studies presented in *Case Studies*)

PART XII

CASE STUDIES

Robert J. Glushko

Overview

We now fulfill the promise of this book, with a set of case study examples that apply the concepts and phases of the roadmap. (The first four case studies appeared in the first print edition of the book. All the others have been contributed by students or other readers of the book and edited for consistency.——Ed.)

Navigating This Chapter

Case Study Template

For the sake of consistency, we employ the questions posed in *Design Decisions in Organizing Systems* as a template for the case studies. We remind you of six groups of design decisions, itemizing the most important dimensions in each group:

- **What is being organized?** What is the scope and scale of the domain? What is the mixture of physical things, digital things, and information about things in the organizing system? Is the organizing system being designed to create a new resource collection, catalog an existing and closed resource collection, or manage a collection in which resources are continually added or deleted? Are the resources unique, or are they interchangeable members of a category? Do they follow a predictable "life cycle" with a "useful life"? Does the organizing system use the interaction resources created through its use, or are these interaction resources extracted and aggregated for use by another organizing system? ("What Is Being Organized?")

- **Why is it being organized?** What interactions or services will be supported, and for whom? Are the uses and users known or unknown? Are the users primarily people or computational processes? Does the organizing system need to satisfy personal, social, or institutional goals? ("Why Is It Being Organized?")

- **How much is it being organized?** What is the extent, granularity, or explicitness of description, classification, or relational structure being imposed? What organizing principles guide the organization? Are all resources organized to the same degree, or is the organization sparse and non-uniform? ("How Much Is It Being Organized?")

- **When is it being organized?** Is the organization imposed on resources when they are created, when they become part of the collection, when interactions occur with them, just in case, just in time, all the time? Is any of this organizing mandated by law or shaped by industry practices or cultural tradition? ("When Is It Being Organized?")

- **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 (or by Whom) Is It Organized?”)
- **Where is it being organized?** Is the resource location constrained by design or by regulation? Are the resources positioned in a static location? Are the resources in transit or in motion? Does their location depend on other parameters, such as time? (“Where is it being Organized?”)

As we discussed in “Where is it being Organized?”, when location is a constraint, it will typically be identified as such in the other questions. As result, we will only examine “Where?” as distinct design dimension in cases where it is warranted.

76. A Multi-generational Photo Collection

Overview. Your grandfather has died, at age 91, and under his bed is a suitcase containing several photo albums with a few hundred photos. Some of them have captions, but many do not. What do you do with them?

Your first thought was to create a digital photo archive of Grandpa's collection so that you and all your relatives could see them, and you would also want to generate accurate captions where none exist. Since you have an extensive digital photo collection of your own in a web-based application, perhaps you can combine the two collections to create a multi-generational photo organizing system.

This project involves digitization, archiving, social media issues, and negotiations with and collecting information from other family members who might have different views about what to do.

What is being organized? It is easy to find advice about how to digitize old photos, but there are more choices than you might think. What resolution and format should you use? Should you do the work yourself or send Grandpa's precious photos to a service and take the risk that they might get lost? Should you do any restoration or enhancement of the photos as part of the digitization process?[1]

More fundamental design questions concern the scope and scale of the organizing system. If you are digitizing Grandpa's photos

1. (Ctein 2010) and (Taylor 2010) are popular guides for photo digitization and restoration.

and combining them with yours, you are skipping a generation. Should not you also include photos from your parents and the rest of Grandpa's children? That generation has both printed photos and digital ones, but it is not as comfortable with computers as you are, and their digital photos are stored less systematically on a variety of CD-ROM, DVDs, flash memory sticks, and SD photo cards, making the digitizing and organizing work more complicated. Do these differences in storage media reflect an intentional arrangement that needs to be preserved? And what about that box full of Super 8 cartridges and VHS tapes with family videos on them, and the audio cassettes with recordings made at long-ago family gatherings?

A family history management system that includes many different resource types is a much bigger project than the one you contemplated when you first opened Grandpa's suitcase. It is easier to consider using separate but related organizing systems for each media type, because there are many web-based applications you could use. In fact, there are far too many choices of web applications for you to consider. You might compare some for their functionality and usability, but given the long expected lifetime of your organizing system there are more critical considerations: whether the site is likely to last as long as your collection and, if it does not, how easily you can export your resources and resource descriptions.[2]

2. For example, http://web.appstorm.net/roundups/media-roundups/top-20-photo-storage-and-sharing-sites/ reviews 20 photo storage and sharing sites and http://photo-book-review.toptenreviews.com/ compares 10 sites for creating printed albums from digital photos in case you want to "round trip" from

Why is it being organized? The overall goal of preserving Grandpa's photos needs no justification, but is preservation the primary goal? Or, rather, is to enable access to the images for far-flung family members? Or is it to create a repository for family photos as they continue to be produced? Alternatively, is it less about the images themselves and perhaps more about collecting family history information contained in the photos, thus making the collection of metadata (accurate information about when and where the photo was taken, who is in it, etc.) most important?

These decisions determine requirements for the interactions that the photo organizing system must support, but the repertoire of interactions is mostly determined by the choice of photo storage and sharing application. Some applications combine photo storage in a cloud-based repository tied to a very powerful set of digital photography tools, but this functionality comes with complexity that would overwhelm your less technology-savvy relatives. They would be happy just to be able to browse and search for photos.

How much is it being organized? Because you realize that a carefully designed set of categories and a controlled tagging vocabulary will enable precise browsing and search, you chose an application that supports grouping and tagging. But not everyone should be allowed to group or tag photos, and maybe some of the more distant relatives can view photos but not add any.

Will your categories and tags include all of those that Grandpa used when he arranged pictures in albums and made notes on the back of many of them? Do you want to allow annotations? Maybe this is a picture from a vacation; if you go back to the same place, do you want to create an association between the pictures?

Grandpa's photos and print photo books for family members.

Do not forget to keep Grandpa's original albums in a safe place, not under a bed somewhere.

When is it being organized? Once you create your categories and tags, you can require people to use them when they add new photos to the collection. Perhaps the existing resource descriptions can be completed or enhanced as a collective activity at a family reunion. Do not put this off too long—the people who can identify Grandpa's sister Gladys, her second husband, and his sister in an uncaptioned photo are getting on in years.

How or by whom is it being organized? You have taken on the role of the editor and curator, but you cannot do everything and having a group of people involved will probably result in more robust organizing. A group can also better handle sticky situations like what to do if people get divorced or have a falling out with other family members; do pictures taken of or by them get deleted?

Other considerations. Maintenance of this collection for an indefinite time raises the important issue or a succession plan for the curator. If only one name is on the account and only that person knows the password, you run the risk of losing access to the photos if that person dies. One of Grandpa's mistakes was dying without clearly specifying his intentions for his photo collection, so whatever you decide you should document carefully and include a continuity plan when you are no longer the curator.[3]

3. (Herbst 2009) is a thoughtful legal primer on the novel property, jurisdiction, and terms of service complexities in gaining access to accounts of deceased people. A popular treatment about what has come to be called the "digital afterlife" is (Carroll and Romano 2011).

77. Knowledge Management for a Small Consulting Firm

Overview. A senior professor who has done part-time consulting for many years is very pleased when his latest book becomes a best-seller and he is inundated with new consulting opportunities. He decides to take a two-year leave of absence from his university to start a small consulting firm with several of his current and former graduate students as his junior consulting partners.

An organizing system for knowledge management is required, but what gets designed will depend on the scoping decision. Is the goal of the system to support the consulting business, or also to support ongoing and future research projects that sooner or later will generate the consulting opportunities?

What is being organized? The professor concludes that since his consulting is based on his research, he needs to include in the new knowledge management system his research articles and the raw and analyzed data that is discussed in the articles. These resources are already organized to a great extent according to the research project that led to their creation. These have been kept in the professor's university office.

The professor also has a separate collection of consulting proposals, client reports, and presentations that he has made at client firms. Because of restrictive university rules about faculty consulting, the professor has always kept these resources in his home office rather than on campus.[1]

1. http://www.spo.berkeley.edu/guide/consultquick.html is an example of such a policy. Indeed, it is because of

In addition to these existing resource types, it will be necessary to create new ones that make systematic and explicit information that the professor has managed in an informal and largely tacit manner. This includes consulting inquiries, information about prospects, and information about specific people in client firms.

Why is it being organized? The professor has usually just done one consulting project at a time, very opportunistically. He has often turned down projects that involved more work than he could do himself. He now sees the opportunity to do much more consulting and to take on more significant projects if he can leverage his expertise in a more efficient way.

The professor can take on the "rainmaker" role to secure new consulting engagements and make the important decisions, and he is confident that he can train and support his new staff of current and former students to do much of the actual consulting work.

The knowledge management system must enable everyone in the firm to access and contribute to project repositories that contain proposals, plans, work in progress, and project deliverables. Much of this work can be reused from one project to another, increasing the productivity of the firm and the quality of its deliverables.[2]

Just as it is essential that the professor's knowledge is systematized and made available via a knowledge management system, so must the knowledge created by the new staff of consultants. The professor cannot expect that all of the students will work for him

rules like these that the professor determined he needed to take a leave of absence from the university.

2. For a high-level theoretical framework about capturing value from knowledge assets see (Teece 1998); for a detailed case study see (Goodwin et al. 2012).

forever, so any knowledge that they acquire and create in the course of their work will be lost to the firm unless it is captured along with the professor's.

The consulting firm probably will not have an indefinite lifetime. After his leave of absence, the professor might return to his university duties, perhaps on a part-time basis. The knowledge management system will enable him to leave the firm in someone else's hands while enabling him to keep tabs on and possibly contribute to ongoing projects. Alternatively, if the firm is doing very well, perhaps the professor will resign his university position and take on the role of growing the firm. A larger consultancy might want to acquire the professor's firm, and the firm's valuation will in part be determined be the extent to which the firm's capabilities and resources are documented in the knowledge management system.

How much is it being organized? A small firm has neither the money nor the people to invest in complex technology and a rigorous process for knowledge management, but appropriate technology is readily available and affordable. Decisions about organizing principles must be made that reflect the mix of consulting projects; resources might be organized in a shared file system by customer type, project type, the lead consultant, or all of these ways using a faceted classification approach.

Standard document templates and style sheets for the resource types created by consultants can be integrated into word processors and spreadsheets. Contact and customer management functionality can be licensed as a hosted application.

Many small teams make good use of wikis for knowledge management because they are very flexible in the amount of structure they impose.[3]

3. (Poole and Grudin 2010).

When is it being organized? The professor's decision to take a leave of absence reflects his belief that getting the firm started quickly is essential if he is to capitalize on his recent bestselling book to generate consulting business. This makes managing the prospect pipeline and the proposal-writing process the highest priority targets for knowledge management.

Much of the other organizing work can emerge as adjuncts to consulting projects if some effort is made to coordinate the organizing across projects.

How or by whom is it being organized? Because many of the early organizing decisions have implications for the types of customers and projects that the firm can take on, only the professor is capable of making most of them. The principal goal of the knowledge management system is to enable the professor to delegate work to his consulting staff, so he needs to enlist them in the design of the organizing system to ensure it is effective.

Other considerations. As the consulting firm grows, it is inevitable that some consultants will be better than others at creating and using knowledge to create customer value, and they will expect to be compensated accordingly. It is essential for the ongoing success of the firm not to let this create disincentives for knowledge capture and sharing between consultants. The solution is to develop a company culture that promotes and rewards them.[4]

4. (Hansen 2009).

78. Smarter Farming in Japan

Overview. Unlike the first two case studies, this is an actual case rather than a hypothetical or composite one. It shares with the first two cases a focus on preserving valuable resources but in the radically different domain of farming.

This case concerns an initiative by Fujitsu, a Japanese technology firm, to apply "smart computing" and lean manufacturing techniques to the agricultural sector, which lags in technology use. Fujitsu is testing a "farm work management system" at six Japanese farms. In this case study we will focus on the farm highlighted in a 2011 Wall Street Journal story.[1]

This test farm is located in southern Japan. It has 60 different crops spread over 100 hectares (about 250 acres), an area slightly larger than the central campus of the University of California at Berkeley.

What is being organized? Sensors are deployed in each of 300 different farm plots to collect readings on temperature, soil, and moisture levels. Video cameras also monitor each plot.

The 72 relatively unskilled workers on the farm are also managed resources. Each of them carries a mobile phone for communication, transmission of pictures, and GPS tracking of their location.

Why is it being organized? The highest-level goal for Fujitsu is to expand its reach as a technology firm by applying the concepts of lean manufacturing, statistical process control, and continual improvement to new domains. Farming is an obvious choice in Japan because it is a relatively unproductive sector where the average age is over sixty. It is essential that farms use more computing

1. (Wakabayashi 2011).

capability to increase efficiency and to capture and reuse the scarce knowledge possessed by aging workers.

The Fujitsu farm work management system supports numerous types of interactions to achieve these goals. For example, workers can send pictures of infected crops for diagnosis by an expert farmer in the farm's office, who can then investigate further by studying recorded video from the affected plot.

As more farms deploy the Fujitsu system, the aggregated knowledge and sensor information can be analyzed to enable economies of scale that will allow separate and widely distributed farms to function as if they were all part of a single large firm with centralized management.[2]

How much is it being organized? The current design of the system treats farm workers as relatively passive resources that are managed very closely. The system generates a daily schedule of planting, maintenance, harvesting, and other activities for each worker. At a daily wrap-up meeting the farm manager reviews each worker's performance based on GPS and sensor readings.

The sensor data is analyzed and organized extensively by Fujitsu computers to make recommendations, both agricultural ones (e.g., what crop grows best in each plot and the work schedule that optimizes quality and yield) and business ones (the profitability of growing this crop on this plot of land).

When is it being organized? The farm work management system is continually organizing and reorganizing what it knows about the farm as it analyzes sensor and production information. In contrast,

2. (Hori, Kawashima, and Yamazaki 2010). Fujitsu expects that the system will eventually integrate business management functions, production history, and operational support for best practices.

the information created by the workers is captured but its analysis is deferred to an expert.

It is conceivable that as the farm workers become more expert as a result of the guidance and instruction they receive from the system that they can be more autonomous and do more analysis and interpretation on their own. It is also likely that the inexorable forces of Moore's law will enable more data collection and more processing of the sensor data at its time of collection, which might result in increased real-time information exchange with the workers.

How or by whom is it being organized? The physical organization of the farm, with 300 small plots of land with 60 different fruits and vegetables, is the legacy arrangement of the farm before the Fujitsu trial began. Because of the sizable investment that Fujitsu has made in the farm to deploy the system, it is likely that the farm manager defers to recommendations made by the system to change crop arrangements. So it is reasonable to conclude that most of the decisions about the organizing system are made by computational processes rather than by people.

Other considerations. Fujitsu built this system for managing farms, but there are several other resource domains with similar challenges about capturing and reusing operational knowledge: vineyards, forests, and fish farms come to mind.[3] It will be interesting to see if the farm work management system can be made more abstract and configurable so that the same system can be used in all of these domains.

Farm crops, vineyards, trees, and fish pens do not move around, so a more challenging application of sensor technologies arises with cattle herd management. Nevertheless, sensors inserted in

3. See(Burrell, Brooke, and Beckwith 2004) for a study of the use of sensor networks in Oregon vineyards.

the genitals of a female dairy cow can trigger a text message to a herd manager's cell phone when the cow is in heat, preventing the economic loss of missing a reproductive cycle.[4]

Somewhat more remote domains for potential application of systems that combine sensor networks with workforce management include sales, field support, and logistics.

4. (Tagliabue 2012). We cannot resist describing this as "sexting" by cows.

79. Single-Source Textbook Publishing

Overview. The fourth case is also an actual case—a self-referential one. It is a case study about the organizing system involved in the creation, production, and distribution of The Discipline of Organizing. See (Glushko 2015).

We have known since the beginning of this project that this book should not just be a conventional text. A printed book is an intellectual snapshot that is already dated in many respects the day it is published. In addition, the pedagogical goal of *The Discipline of Organizing* as a textbook for information schools and similar programs is made more difficult by the relentless growth of computing capability and the resulting technology innovation in our information-intensive economy and culture. We think that the emergence of ebook publishing opens up innovative possibilities as long as we can use a single set of source files to produce and update the print and digital versions of this book.

What is being organized? The content of this book began in early 2010 as more than 1000 slides and associated instructor notes for a graduate course "Information Organizing and Retrieval" that Robert J. Glushko, the primary author and editor of *The Discipline of Organizing*, was teaching at the University of California, Berkeley. These slides and notes were created in XML and transformed to HTML for presentation in a web browser.[1]

1. (Wilde and Catin 2007). Looking back it seems ironic to start with a single-source XML publishing system, abandon it to author the book in Word, and then convert

The first decision to be made about resource organization led to the iterative sorting of the slides from 26 lectures into the 10 chapters in the initial outline for the book. The second decision concerned the granularity of the new content resources being created for the book. The team of authors was organized by chapters, which made chapters the natural granularity for file management and version control. Because authors were widely dispersed we relied on the Dropbox cloud storage service to synchronize work. Nevertheless, the broad and deep topical coverage of the book meant that chapters had substantial internal structure (four levels of headings in some places), and many of these subsections became separately identified resources that moved from chapter to chapter until they found their natural home.

In addition to the text content and illustrations that make up the printed text, we needed to organize short videos, interactive examples, and other applications to incorporate in digital versions of the book.

Finally, it has been essential to view the software that transforms, assembles, formats, and assigns styles when turning source files into deliverable artifacts as resources that must be managed. For the first and second editions of the book, we were fortunate to get much of the software required to build both print and ebooks from O'Reilly and Associates, an innovative technology publisher that has been developing a single-source publishing system called Atlas. Because we have recently been experimenting with including richer interactivity and navigation capability, reader-controlled personalization, and other features that go beyond what Atlas enables, we now use our own custom-built single-source publishing system.

the files Word back to XML to enable single-source publishing.

Why is it being organized? Publishing print and ebook versions of a text from the same source files is the only way to produce both in a cost-effective and maintainable fashion. Approaches that require any “hand-crafting” would make it impossible to revise the book on a timely schedule. Furthermore, a survey of Berkeley students in the summer of 2012 revealed a great diversity of preferred platforms for reading digital books that included laptop computers, Apple and Android tablets, and seven different dedicated ebook readers. Only an automated single-source publishing strategy could produce all these outputs.

The highly granular structure for the content resources that comprise this book makes cross-referencing vastly more precise, making it easier to use the book as a textbook and job aid. It will also make it easier to maintain and adapt the text for use in online courses. (The emerging best practice for online courses is to break up lectures and study content into smaller units than used in traditional classroom lectures.)

How much is it being organized? The nature and extent of resource organization for this book reflects its purpose of bringing together multiple disciplines that recognize organizing as a fundamental issue but from different perspectives. The book contains many specialized topics and domain-specific examples that might overwhelm the shared concepts. Our solution was to write a lean core text and to move much of the disciplinary and domain-specific content into tagged endnotes. These categories of endnotes are somewhat arbitrary, but the authoring task of identifying content to go into endnotes is a non-trivial one.

The extent of resource organization is also affected by the choice of XML vocabulary, and we carefully considered whether to choose DITA or DocBook. DITA has the benefit of having more native support for modular authoring and transparent customization and updating, but DocBook is much older and hence has better toolkits. We eventually chose DocBook.[2]

When is it being organized? Despite the fact that the lecture notes with which the book began were in XML, we decided to author the book using Microsoft Word. Many of the authors had little experience with XML editors, and the highly developed commenting and revision management facilities in Word proved very useful. This tradeoff imposed the burden of converting files to XML during the production process, but only two of the authors were still working on the book at that stage, and both have decades of experience with hypertext markup languages.

How or by whom is it being organized? The chapter authors used Word style sheets in a careful manner, tagging text with styles rather than using formatting overrides. This enabled a conversion vendor to convert most of the book from Word to XML semi-automatically. Some cleanup of the markup is inevitable because of the ambiguity created when the source markup with Word styles is less granular than the target markup in XML. We do not know whether the amount of work left for us was atypical.

Nevertheless, waiting until the book was substantially finished to convert to XML meant that we were also deferring the effort to mark up the text with cross references, citations, glossary terms, and index entries, because these types of content were not included in the Word authoring templates and style sheets. As a result, a substantial amount of effort has been required of our copy and markup editor that could have been done by chapter editors if they had authored natively in XML. However, having a single markup editor has given this book a more consistent and complete

2. (Kimber 2012) seems destined to become the definitive resource for DITA-based publishing. The definitive source for DocBook has long been (Walsh 2010).

bibliography, glossary, and index than would be have possible with multiple authors.

Other considerations. Because every bit of content in the book is tagged as either "core" or discipline-specific, our source files collectively represent a "family of books" with 2048 different members, any one of which we can build by filtering the content to include any combination from zero to eleven disciplines. It is impractical to publish this many editions, but we hope to use this flexibility to enable instructors to tailor the text for a wide range of courses in many different academic disciplines and customize the text for both graduate and undergraduate students. Better still would be an approach that defers the generation of a particular version of an ebook from "publishing time" to "reading time." The same algorithms apply, but now the reader decides when and how to apply them, enabling the dynamic configuration of the book's content. This radical capability is experimental as of August 2015, but we expect it to generally available before too long.

This design for a book challenges conventional definitions of book editions and forces us to imagine new ways to acknowledge collaborative authorship. But asking "What is *The Discipline of Organizing*?," given these new authoring and publishing models, is a similar question to the one asked in *Resources in Organizing Systems*, "What is *Macbeth*?"

80. Organizing a Kitchen

By Emilie Hardman, April 2013.

Overview. Just about everyone has a kitchen in their home or apartment, and most kitchens contain many of the same resources. These include pots and pans, dishes, bowls, drinking glasses, silverware, and cooking tools of various kinds. Kitchens are also often the location for organizing food items, cooking ingredients, spices, wine, and other beverages. Kitchens also invariably contain refrigerators and freezers for storing prepared and preserved food.

The organizing system for a kitchen is highly influenced by the size, shape, and arrangement of the counters, cabinets, shelves, and other parts of the physical environment of the kitchen. A person building a new home might be able to design this kitchen environment, but most people treat this as a given and work within its affordances, often because there are limits to how much the physical environment can be easily changed.

Kitchen Organizing System

My kitchen. I did my annual deep kitchen clean and it deserved a picture.

(Photo by Emilie Hardman. Creative Commons CC-BY-SA-2.0 license.)

What is being organized? Our wine, wine glasses, cocktail glasses and ingredients, as well as tea and coffee stuff were stored in the cabinet by the fridge, close to the center worktable so people could have easy access to them. Because of space limitations, this meant that our water glasses had to be somewhere else, but as we would usually put out water for dinner parties or have a pitcher and glasses on a tray when people came over, we thought this was reasonable, since the things people would most often be looking for and need easy access to for themselves would be these more social drinking glasses.

We also bought a freestanding worktable with a butcher's block and stainless steel for pastry and chocolate work, as well as extra counter space in general. It worked as a prep space and as an area to lay out finished dishes or drinks for people to serve themselves when we had parties.

Some kitchen tools were kept with the food items to which they applied: for example, the coffee and the coffee grinder, or the cutting board, toaster, bread knife and bread all together. Other tools were kept with like tools: potato peelers, julienne tools, knives, etc. This was probably because of the kind of flexibility something like a potato peeler would have versus a coffee grinder; it also made more sense to put lots of these little things together in a drawer rather than leave them strewn out around the apples or potatoes.

Pots and pans had their own spaces and were stacked within one another; same with dishes. Most frequently used things were given preference over specialty tools.

Other things that were organized around the social dimensions of the kitchen were some food items and serving elements. For example, we used bowls to organize chocolate bars and treats that might easily be grabbed to set out and serve. Similarly, we kept stacks of serving bowls easily at hand so we could empty pretzels or tortilla chips, olives, etc., quickly and casually.

Why is it being organized? We wanted to emphasize a feeling of comfort and openness in our kitchen, so people would feel free to get what they wanted when they needed it. It also had to work on a practical level to be an efficient work area in a small space, so those concerns had to be balanced as well.

When is it being organized? We ended up moving silverware at one point because friends would consistently open a particular drawer in our center work island to look for silverware. Initially, I had specialized tools in that drawer because they were what I would reach for when I was working on something like making chocolates, but because of the continuous confusion, we moved those tools to another drawer and put the silverware where people seemed to expect it.

The fridge and freezer was organized by type of food for orderliness, ease of access, and immediacy of knowing when we had. We have a pull-out freezer, so things could get a little hidden, but assuming no one had compromised the system, you would know it was frozen fruit all of the way down in one segment and flours in another.

Some food items demanded different placement or storage based on their ripeness, the season, etc. In August we might be overrun with tomatoes, for example, and the window sills would fill up with them, whereas we would usually put them in a bowl if there were just a few.

How or by whom is it being organized? I think one thing to sum up would be to say that my partner is a librarian and I am trained as an archivist. We both care about classification and public service, so as people who also entertain a lot, I think these very practical and intuitive systems of grouping things is a motivation.

My father, an engineer who in his retirement does a lot of woodworking, built two cabinets that would *just* fit into the space and provide more storage than the two upper cabinets and three base cabinets provided in the kitchen.

Other considerations. The whole kitchen was not organized around guests, though. We also arranged things to be practical for cooking and for space saving. Food in the cabinets was organized by general function: for example, there was a shelf of dried beans in jars, another of dried chilies and spices—things that give flavor. Spices were organized within that by general type in rows and then alphabetically within those rows. This was because the rows helped group things which might be likely used together (e.g., cinnamon, cloves, mace, nutmeg) and alphabetically because so many of them look the same from the outside; knowing that the oregano would necessarily be shelved before the thyme was useful. Beans, though, because they are more immediately identifiable, less used, and certainly not as often used in concert (as one would with spices), I was a little more loose with and sometimes just arranged to a general aesthetic preference; if we had heirloom money beans, I might have preferred to see them over the standard red lentils, for example.

81. Earth Orbiting Satellites

By Daniel Brenners, December 2014.

Overview. Twenty two thousand miles above our heads, a global race for orbital real estate is underway. A single circular orbit around the Earth, called the geostationary Earth orbit (GEO), is the only area in space that allows a satellite to remain in a fixed point in the sky above Earth's surface while it rotates.[1] This prime location allows for satellites to have consistent communication with the ground below. Satellite television, a $100 billion industry, relies on satellites within the GEO to broadcast signals to homes across the world. Global positioning systems (GPS) and military applications also depend on satellites within this thin ring around the Earth. Unfortunately, space is severely limited in the GEO, and tension is growing over who gets to send their satellites to this valuable parking lot in the sky. The principles used to organize which satellites get to be placed in the GEO have many unforeseen legal and sociopolitical complications. As room becomes limited, it becomes increasingly important to find a solution to the problem of multiple organizing agents competing to organize this system to support varying interactions.

What is being organized? The scope of resources being organized are the satellites being deployed to the GEO. These are physical objects that have been launched into orbit. The satellites are each unique and are able to provide a variety of interactions. The only unifying attribute that they share is that they are computers that are able to send and receive radio signals to and from Earth. To stay in orbit, they are also able to adjust their position with propulsion systems.

This organizing system is designed to manage a collection in which resources are continually added and removed. The International Telecommunications Union (ITU) records which portions of the orbit are already occupied.[2] Satellites cannot stay in the orbit forever, as they expend lots of energy performing computational processes and maintaining orbit, and eventually run out of power. The resources follow a lifecycle that is unique to each individual resource, but the timescale is typically one to fifteen years.[3]

Why is it being organized? Satellites are being organized in the GEO to support several interactions. The GEO allows satellites to move at the same rate as the Earth, giving it a stationary view of more than 40 percent of the Earth's surface. Such a view is ideal for broadcasting signals to large regions and performing remote sensing, such as weather forecasting. They also serve as crucial relay points to transfer telecommunications across the globe. Other interactions that these satellites provide include surveillance, scientific research, global positioning, navigation, and military reconnaissance.[3] Longitudinal positioning along the GEO shapes which interactions can occur and which users can interact with the satellite. For instance, a satellite directly over the Atlantic Ocean may not be well suited to broadcast a television signal, but may be positioned to relay signals from North America to Europe.

The users are practically everyone on Earth. Civilians use geostationary satellites directly when they use GPS or need to have a call relayed to distant regions of the world. Commercial organizations, such as television providers, use these satellites to broadcast signals down to viewers. Geostationary satellites are also particularly useful for early warning systems used by the military to detect ballistic events around the globe.

How much is it being organized? If resources are able to be placed in the GEO, they are placed in a vacant slot that the applicant chooses, based on what types of interactions they want to support and what users they want interacting with the satellite. To prevent signal interference and collision, satellites need to be placed very far apart, leaving only 2,000 total orbital slots where satellites can be placed in the GEO.[4] The ITU uses a first-come, first-served organizing principle to decide which resources are placed into orbital slots, provided the applicant completes the lengthy application process.

The organization applying for the slot chooses where to place its satellite. The ITU catalogs these slots as degrees longitude, and includes other resource descriptions such as the name of the satellite, country of operator, types of users, mass, expected lifetime, and contractor.[3] Organizations choose to place satellites around the longitude of the Earth that the satellite is supposed to interact with. Since the latitude is fixed at zero degrees, countries with the same longitude but different latitudes (countries directly north or south of each other) must vie for the same slots.

When is it being organized? Satellites are added as soon as they can be approved by the ITU and launched into orbit. At the end of their life cycle, the Federal Communications Commission mandates that U.S. satellites are pushed into what is called the graveyard orbit, which is a few hundred kilometers outside of the GEO.[5] At this point, another satellite can be added to the vacant slot via the ITU application process.

How or by whom is it being organized? Many organizing agents are competing with each other to organize this system according to their own needs. Applications to occupy the GEO come from countries, scientific organizations, companies, and civilians. Satellite TV companies such as DirecTV, Dish Network, and Intelsat own a large number of the slots across the western hemisphere. Countries such as the United States, Russia, and the United Kingdom own a majority of the military satellites, and multinational European organizations own a large share of orbital slots as well.[3]

Other considerations. Although the ITU serves as an authoritative entity for this organizing system, the reality is that the ambiguous legality of ownership in outer space means that anyone can attempt to organize this system. The ITU is in place to perform the useful task of cataloging occupied slots and facilitating the filling of vacancies, but it has no way of enforcing these guidelines.

This organizing system is interesting because many agents are attempting to organize the same system. There are also interesting social implications that stem from the system's principles of organization. The first-come, first-served system of the ITU has the effect of allowing only technologically advanced organizations to manage the collection. It does not take into consideration that by the time many countries are finally ready to use this type of technology, there will be no more room in the GEO belt.

Ironically, the only legal claim to sovereignty that has been made of this organizing system has been from countries that, generally, have no means of organizing it themselves. In 1976 eight equatorial countries, which lie directly below the GEO belt, stated that they had exclusive rights over these slots in a document known as the Bogotá Declaration.[6] The tenuous claim was that the orbit is not a part of outer space, because its existence is solely dependent on Earth's gravity, and that the earth within the borders of the equatorial countries creates GEO with its gravitational pull. Many experts disagree, stating that the gravitational pull from the moon and other celestial bodies defines the GEO, and state that the orbit does indeed lie in outer space because it is further than 100 kilometers from Earth. This demarcation, known as the Kármán line, is a widely accepted definition of when space begins.[7] This would then make the GEO fall within the 1967 Outer Space Treaty, effectively leaving no possibility for ownership of the orbit.

Finding a dividing line between space and Earth's atmosphere is an interesting topic, especially considering that ownership of valuable resources may be decided based on what is included in the category of space versus the category of atmosphere. In this case, the Kármán line roughly represents the altitude at which an aircraft would have to propel itself faster than the speed at which the Earth rotates to establish enough lift to keep itself up. While this is not intuitive (hardly carving nature at its joints), it does serve as a useful demarcation that is not completely arbitrary. It can be seen as a goal-based category, where the goal is using traditional means of traveling through the air using aeronautics. It makes sense that this is the line the Fédération Aéronautique Internationale uses to divide astronautics and aeronautics.

The limited availability of spots in the GEO, along with the relatively small number of countries able to launch satellites, has the potential to further divide countries. By the time most countries will be able to launch satellites, there will likely not be any room left. Although there are only around 400 satellites currently in geostationary orbit, there are already more filings for ITU applications than there are spots available.[4] Only a select few countries will be able to take advantage of the GEO, leaving others to depend on these countries for communication, scientific research, and surveillance. Furthermore, this could limit the interactions of these less developed countries to those interactions dictated by the countries with geostationary satellites. In particular, these developed countries can greatly influence the information that citizens in other countries can receive via satellite.

But even within the technologically advanced countries, competition for orbital slots may be heating up. In early 2014, the US unveiled its Geosynchronous Space Situational Awareness Program (GSSAP), which aims to create maneuverable satellites that monitor and protect the precious GEO belt.[8] This reveal comes only months after China was seen practicing its anti-satellite missile capabilities.[9] In Russia, $300 million is being spent to construct a craft that would act as a "space broom" to push satellites out of geostationary orbit. The US has a similar program, named the Phoenix project under DARPA, developing a robotic device that can help maintain satellites and possibly dismantle others without causing excess space debris.

Although this might simply be countries attempting to flex their military muscles, these technologies represent a newfound ability for countries to organize resources in the GEO to fit their own agenda. Years ago, the countries that were able to get satellites into orbit were the ones that could reap the benefits. Now, it seems that we may be entering an age where a country's ability to make room for itself, possibly by force, will determine if it can make use of precious interactions created by these limited resources.

Notes: The following notes relate to this case study.

1. NASA Jet Propulsion Laboratory Basics of Space Flight Section 1 Chapter 5: Planetary Orbits http://www2.jpl.nasa.gov/basics/bsf5-1.php

2. ITU Space Services Department (SSD) 2014 http://www.itu.int/ITU-R/go/space/en

3. Union of Concerned Scientists Satellite Database http://www.ucsusa.org/nuclear_weapons_and_global_security/solutions/space-weapons/ucs-satellite-database.html#.VJKNXmTF-5I

4. Posen M., Have We Got a Slot? RPC Telecommunications Ltd. World Space Forum Dubai March 2010 http://www.rpctelecom.com/files/Have We Got A Slot.pdf

5. De Selding P., FCC Enters Orbital Debris Debate. Space News, 28 Jun. 2004

6. Finch M., Limited Space: Allocating the Geostationary Orbit. Northwestern Journal of International Law Vol 7 Issue 4 Fall 1986

7. Haraszti G., Questions of International Law Volume 2. Akademiai Kiado Budapest 1981

8. Hsu J., Global Conflict Could Threaten Geostationary Satellites: China, Russia and the U.S. have the ability to destroy one another's eyes in the sky. Scientific American March 31, 2014

http://www.scientificamerican.com/article/global-conflict-could-threaten-geostationary-satellites/

9. Shalal-Esa A. U.S. sees China launch as test of anti-satellite muscle. Reuters May 2013 http://www.reuters.com/article/2013/05/15/us-china-launch-idUSBRE94E07D20130515

82. CalBug and its Search Interface Redesign

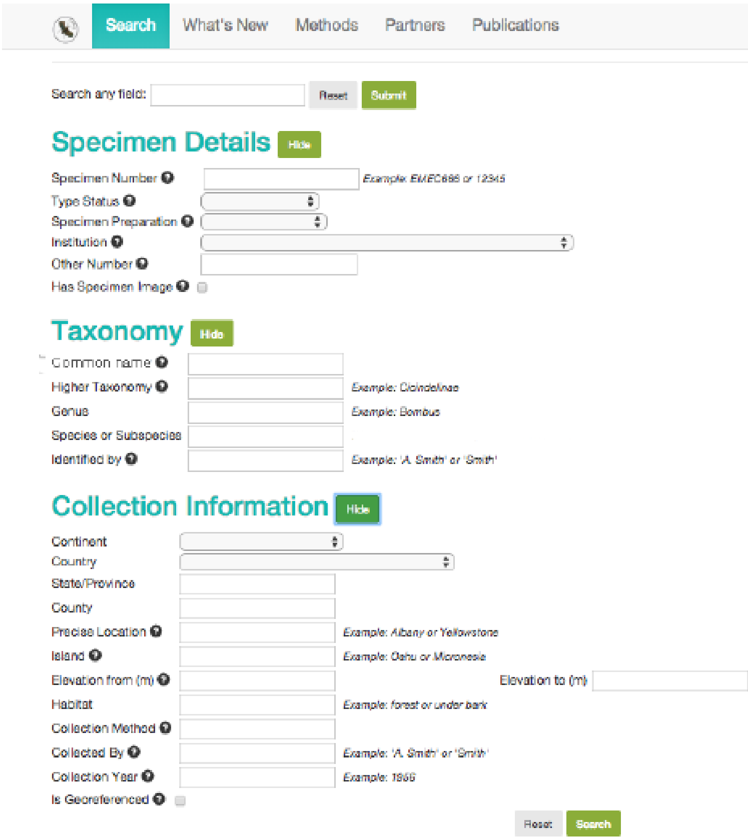
By Gracen Brilmyer, December 2014.

Overview. The CalBug project, housed out of the The Essig Museum of Entomology at the University of California, Berkeley, is a collaborative initiative between nine California institutions with a goal to digitize over a million specimens. Digitization involves imaging both specimens and their labels as well as storing their collection info in a database. The CalBug project also is attempting to georeference, or locate the original latitude and longitude coordinates, for these million specimens (some dating back to the 18th century) so that they can be better used for research. The project uses many student workers, graduate students, and volunteers to capture the images and data. Over the past few years, it has participated in the Notes from Nature project, which helps connect citizen scientists to scientific research. Through the images generated of the specimen labels by the team at the Essig Museum, citizen scientists digitally transcribe the data that can be read from the image. The Essig, after each label is transcribed by 24 citizen scientists, runs an R program to find the most accurate transcription and transfer it into the Essig's database. These combined efforts have accumulated in over 209,000 specimen records and over 400,000 images and counting. This project has a large scope and an ever-increasing scale.

What is being organized? The insect specimens in the CalBug project are digitized on an individual level, with unique identifying numbers, and new specimen records and their associated data are continually being added to the digital collection. Both the specimens and their data are being organized. Existing groups of specimens are prioritized for digitization and new physical specimens are accessioned into the collection and are databased upon arrival.

Why is it being organized? An individual specimen's associated data can be highly variable; however, as long as a specimen has the time and place of its collection (no matter how vague) associated with it, it is valuable research material. The physical specimens are organized to facilitate the collection manager's use of the collection. When physical specimens need to be borrowed, they must be efficiently found, packaged, and sent out on loan, so fastidious organization is key when locating thousands of specimens. The digital organization of the collection also facilitates the duties of museum staff and the collection manager by allowing for expanded interaction with the collection by using the database. The digital collection's web interface, undergoing a redesign as of the time of this writing, makes the collection accessible for researchers and novices alike, as well as to foster data sharing to other data repositories. Since the specimen data follows digital curatorial standards, a web interface that allows these fields to be easily searchable and navigable can add to the use of the collection for a broader audience, which is a major impetus for the redesign.

CalBug Search Interface



CalBug's redesigned web search interface

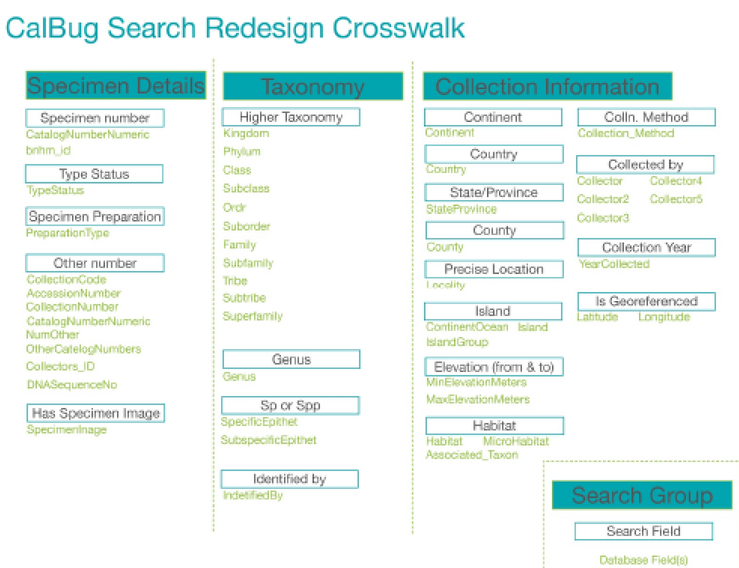
How much is it being organized? As discussed in the previous section, the specimens and their information are subject to multiple levels of organization, and each level of organization supports a different type of user. The data of the CalBug Project is organized according to Darwin Core (DwC), a standard "designed to facilitate the exchange of information about the geographic occurrence of species and the existence of specimens in collections."[1] Certain specimen attributes have concrete institutional parameters, such as unique identifying numbers and taxonomic identification, while others have less strict parameters (e.g. a precise location of where a specimen is found), although they still must use specific DwC fields. Although there are institutional taxonomies in place for information associated with a specimen's collection and identification, the CalBug search interface design in Figure: CalBug search interface allows for an outward-facing reorganization of the existing fields.

When is it being organized, and by whom? The categorization and organization happens at multiple times for one specimen. If identified, the specimen is already inserted into the taxonomic classification scheme—the hierarchy of how species are related. This scientific warrant is inherited and replicated in the physical curation of the collection, and specimens are further sorted (within a taxon) by geographic region. Aligning with taxonomic categories provides a clear hierarchy for sorting and locating physical specimens and, with changes in taxonomy having to be published, makes collection maintenance fairly consistent.

The specimens are organized a second time when they are databased, either by interns or through Notes from Nature. The data is stored in a MySQL database that uses mostly DwC fields, an institutional taxonomy for specimen data. The digitization of specimens, through utilizing DwC institutional semantics, makes collection maintenance, governance, and interaction easier, as the collection manager can search in a multifaceted manner, better understand the holdings of the museum, and track specimens for loans. The unique specimen numbers allow for individual tracking, and the other DwC fields provide multiple areas for accurate search and retrieval.

For the CalBug web search interface, the specimens retain their classification hierarchy within the database. However, the outward-facing search fields aim to serve a broader audience, not just the collection manager and museum staff. Thus the search application organizes the resources a third time "on the way out" of the database in response to a user query. As this design is optimized for researchers and students, the classification appears to focus more on taskonomy instead of the institutional taxonomy (see [Figure: CalBug search interface](#)). The 20 search fields provided in the search interface, while actually searching through the ~100 fields in the database, facilitate precise information retrieval. Although fewer search fields might yield lower accuracy, user testing has shown that the new search design improves accuracy by not requiring users to know exactly which DwC field to query.

Crosswalk Table



This crosswalk table maps the fields in the CalBug search interface to the underlying database columns.

The search is further expanded by having a ‘Search any field’ box, which literally looks in every DwC field for a term, as well as a

"Common Name" field, to support novice searches, such as "beetle" and "butterfly" instead of "coleoptera" and "lepidoptera." The intrinsic properties of the specimens lend the results to simple (alphabetic and numeric) sorting as well as filtering (through the "Refine" option) on the list view of the results pages. Additional views of results, including a map view showing collection locations and a grid view that displays specimen photos, help users locate desired specimens and reorganize as needed to suit their needs.

Notes: 1. http://wiki.tdwg.org/twiki/bin/view/DarwinCore/WebHome

83. Weekly Newspaper

By Ian MacFarland, December 2013.

Overview. A weekly neighborhood newspaper in New York City now covers the entire borough of Queens. Rather than publish a single weekly edition for this highly diverse area of more than 2 million people, its owners have opted to produce 14 separate editions, each centered on a different neighborhood. All editions share a deadline, delivery schedule, and staff pool, but each has unique content tailored to its target readers.

What is being organized? The newspaper's resources—its content—consist mainly of articles and photos generated by staff and freelance contributors throughout the week. Often, newspapers will assign their reporters to beats based on subject matter (politics, education, "cops and courts," etc.), making them domain experts who cover stories on that beat throughout a wide geographical area. However, because of this paper's historical orientation toward "hyper-local" neighborhood news, it has given each of its seven full-time reporters a more granular geographical beat that corresponds to two of the 14 editions' coverage areas, within which they are responsible for general assignment reporting. Most reporters also have a specialty for covering news that is of more general interest throughout the borough, such as citywide government or transportation issues, and they will include coverage of these domains in their story budgets for the week as well. The staff maintains a centralized story list that includes a handful of resource descriptions for each story: its slug (an abbreviated, descriptive name, including tags for its relevant neighborhoods), its length, and whether it has "art."

Why is it being organized? The media market in New York is crowded and extremely competitive, and this newspaper believes its competitive edge lies in its laser-focus on individual neighborhoods. Furthermore, most of its readers are subscribers who receive the paper in the mail, not newsstand buyers. As a result, the paper generally eschews the familiar tabloid approach of splashing the most salacious story of the week across the front page and usually fronts two stories that are "small-bore" but extremely relevant to the neighborhood, such as the doings of local school or government officials, notable crimes, or human-interest stories featuring neighborhood residents. The deeper into the paper one goes, the less local its content becomes, and stories often appear in more than one edition, in different locations and even with different headlines, to tailor them to an appropriate level of localization.

On a more general level, of course, the paper must support the conventional interactions all readers expect from newspapers. Readers are rarely expected to progress through the paper from front to back, so it supports a wide variety of reading styles; large headlines and photos and concise, compelling story "ledes" (opening paragraphs) facilitate skimming and scanning interactions, and dividing the paper into sections, such as "Opinion," "Sports," and "Arts & Entertainment," lets readers skip directly to their areas of interest after turning past page one.

How much is it being organized? The level of organization behind the scenes at this small, local newspaper is surprisingly complex. The primary organizing principle that determines a story's placement is its relevance, which is a function of location granularity (does it directly affect the people of this neighborhood? Did it happen here?), significance (will readers find it important?), and time (is it old news? Has anyone else reported it yet?). Counterbalancing that is the economic reality of the struggling newspaper industry, which results in often severely limited space for the news (because paper and press time are costly physical constraints) and manpower with which to produce all 14 editions before deadline. The result is a hierarchical system in which the 14 editions are categorized into three zones; in each zone, about two-thirds of the pages are common to all editions, and the remaining third (including, most crucially, pages one through three) are unique to each single edition. Thus, for instance, a general-interest story about transportation need not be laid out 14 separate times, but one about a fatal car accident can appear on page one for the neighborhoods where it occurred and where the victims were from, and further back (or not at all) for other neighborhoods.

When is it being organized? In a weekly news cycle, selection, creation, and organizing of editorial resources is largely concurrent. The story list is updated on a rolling basis throughout the week, and an article or photo's placement in the paper is often planned based on its intended subject matter well in advance of when the resource is actually created. However, organizing must be completed long before it reaches its intended users, because the final layouts must be printed, collated, and mailed to readers, which, due to logistical concerns, takes several days—so the paper is laid out on Tuesday (as late as possible to maximize the window for ad sales), printed on Wednesday, and delivered by the Postal Service on Thursday or Friday.

How or by whom is it being organized? Human agents—specifically, editors—are the newspaper's primary organizers. They rely heavily on the judgment of the reporters, who are most familiar with their beats, to determine a story's relevance and placement for each edition, as well as their own news judgment, assessment of the story's quality, and estimation of where the story will physically fit based on ad placements (which are decided first). The implementation of their organizing system is carried out by page layout designers, with some software automation on the part of the paper's content management system.

Other considerations. Part of the grind of a weekly news cycle is that the effectivity of the paper's resources is never guaranteed; when the next edition comes out, they all become yesterday's news, and one never knows when new developments will render a story irrelevant or incorrect; in fact, because of the latency between layout and delivery, a story's effectivity may even expire before its publication.

84. The CODIS DNA Database

By Becca Stanger, December 2013.

Overview. Operating on a local, state, and federal level, the Combined DNA Index System(CODIS) is the FBI DNA database. As of October 2013, the National DNA Index(NDIS), or the federal level of the CODIS, contained over 10,647,800 offender profiles, 1,677,100 arrestee profiles, and 522,200 forensic profiles. Designed to help solve crimes, this database has generated over 255,400 hits and has aided over 216,200 investigations. While this organizing system has played a crucial role in reducing crime by enabling more interactions in the law enforcement agency than ever before, it provokes numerous ethical questions worth exploring.

What is being organized? The CODIS database maintains digital records or "DNA profiles" for a wide range of people involved in criminal justice cases, including convicted offenders, arrestees, missing persons, and more. Specifically, these profiles are measurements of one or two alleles of 13 predetermined unique genetic sequence loci. These precise measurements provide enough granularity for the profiles to uniquely identify a single individual.

These resource descriptions are generated, often with polymerase chain reaction technology, from the original DNA specimen resources by accredited laboratories nationwide. Upon creation, the resources themselves—the specimens—are kept at the laboratories, while the resource descriptions—the digital profiles—are added to the CODIS database. No offender personal identifiers are assigned to the profiles; however, information on the submitting agency, specimen, and personnel is stored with the profile.

Rather than focusing on collecting resource descriptions, the FBI could have chosen to collect the original resources themselves.

Presumably, though, this level of coordination of physical resources (e.g., shipping, storage, maintenance, etc.) would have placed an additional cost on the federal government and required legislative approval. Thus, it is understandable that the FBI would choose to minimize cost and effort by focusing on the resource descriptions alone.

Why is it being organized? In the past, law enforcement agencies were limited to solving crimes within their geographic region. A detective working on a murder in California, for example, may never have heard of a related murder in New York. The CODIS database organizing system encourages that coordination between law enforcement agencies in an effort to solve crimes.

With 10,647,800 offender profiles in the NDIS alone, though, the massive CODIS database required an organizing system in order to prove useful to the law enforcement agencies involved. The successful creation and maintenance of this organizing system has offered newfound interactions to a wide variety of government officials. In addition to law enforcement agencies, judicial courts, criminal defense agencies, and population statistics agencies can access the CODIS organizing system, enabling them to perform a wide variety of functions, including identifying potential suspects in criminal investigations, identifying missing persons, collecting population statistics, and exonerating convicted criminals.

How much is it being organized? As mentioned previously, the high degree of resource description granularity in measuring 13 specific genetic sequence loci enables DNA profiles to uniquely identify each individual in the database. That being said, the DNA profiles are not simply heaped into one massive database.

Instead, the databases are maintained on both a state and federal level. A new profile might be checked against a smaller state database as well as the larger national one. In addition, the databases are divided into different indices dependent on the DNA source, including an offender index, arrestee index, forensic index, and missing persons index.

This division of the database into separate indices poses a tradeoff dilemma, though. If CODIS did not subdivide the database into federal, state and source indices, it is possible the algorithm would be able to find more obscure hits, since the search parameters would be broadened. This increase in hit frequency might result in more investigations aided.

The tradeoff, however, is that the broadened search parameters would also necessitate a more complex search algorithm and a longer search process. This delay would most likely lead to fewer hits overall. Thus, in government institutions where time and resources are limited, it is more important for the CODIS organizing system users to generate a larger number of hits with subdivided databases than more accurate hits in one collective database. Categories in the CODIS organizing system help simplify the interaction processes.

When is it being organized? DNA profiles enter the CODIS organizing system when participating, accredited local, state, and federal laboratories submit them. Thus, the laboratory technicians handling the resource and resource description decide on a case-by-case basis how a given profile should be categorized and which indices it should be added to and checked against.

That being said, the lab technicians are given strict standards on how a given DNA profile should be categorized. These standards vary state by state depending on state law.

How or by whom is it being organized? Beyond laboratory and state involvement in CODIS, the FBI ultimately maintains and oversees the CODIS database. It maintains the software and search algorithms, performs searches throughout the system, and oversees strict quality assurance standards for all participating laboratories.

To avoid the risk of bias or error amongst lab technicians, the FBI could potentially choose to instead perform the laboratory processing and categorization themselves. This alteration, however, would present new challenges, such as new federal costs related to maintaining and processing the resources mentioned previously. In addition, pulling together all resources into a FBI processing center would necessitate a meticulous record of the resource's originating state to ensure resource descriptions are categorized in accordance with state laws. The FBI's strict maintenance of standards and laws is the best option for addressing the risk of error and bias.

Other considerations. The CODIS organizing system presents a wide range of intriguing ethical questions surrounding race, gender, criminal justice, and privacy. Perhaps the most hotly debated issue surrounding DNA databases arose when the private DNA testing company 23andMe announced that it would discontinue the sale of its genetic tests in response to FDA demands, prompting more media questions than ever before on the maintenance and use of DNA databases.

Likewise, many have questioned the legitimacy of the CODIS maintenance of DNA profiles. The ACLU, for example, has noted the possibility of "function creep" in the maintenance of a government DNA database which could lead our country down a slippery slope towards a "brave new world" where private genetic information could be collected and used in abusive, discriminating manners.

With the commercial surveillance of 23andMe and government surveillance by the NSA at the forefront of media attention, it is possible we will see more attention turned to the legitimacy of the maintenance of the CODIS organizing system in the coming years.

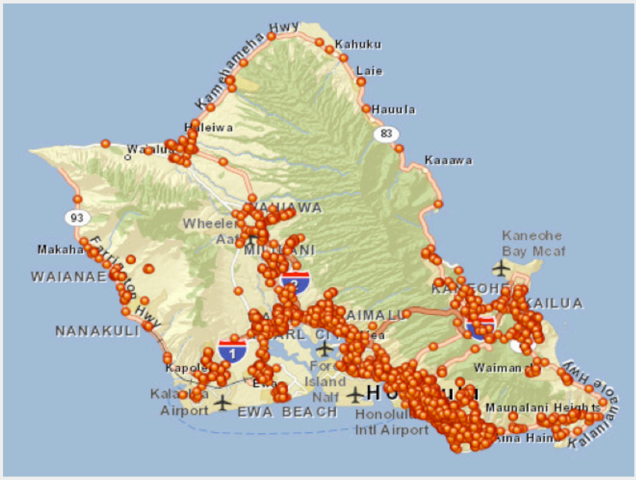
85. Honolulu Rail Transit

By Carlo Liquido, December 2015

Overview. The Honolulu Rail Transit Project is an urban rail rapid transit system under construction in Honolulu on the island of O'ahu, Hawaii. Honolulu's notoriously bad traffic has plagued locals and tourists for decades, and for almost as long, proposals to address the traffic problems and pay for the solution have been very contentious and political. Construction began in 2011 and is expected to finish in 2019, but delays have been frequent.

What resources are being used? The new railway transit system under construction in O'ahu will run along the southwest region of the island spanning a total of 20 miles, from East Kapolei to Downtown Honolulu with a total of 21 stops strategically placed throughout. There are a number of ways in which one could scope this project. What are the cultural and political limitations? What are the environmental effects and resources that will be indirectly affected? What are the topographic constraints of a railway system in Hawaii? In terms of the scope of my analysis, however, the people—namely the things the organizing system is intended for—are the primary resources. The principle guiding the organizing system is to reduce traffic and make the traveling experience more efficient as a whole.

O'ahu Traffic



O’ahu traffic is usually congested, especially in and around Honolulu on the south and southeast sides of the island

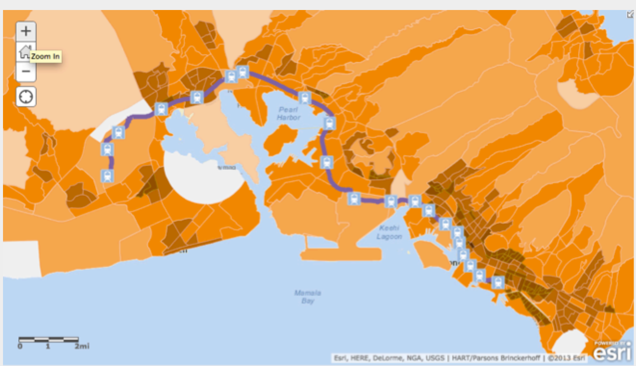
(Hawaii Dep. of Economic Development and Tourism. CC-BY-2.0.)

Why are the resources organized? The guiding principle behind the organizing system of a rail transit system is to reduce traffic and make commuting more efficient. According to the Department of Business, Economic Development and Tourism, the amount of traffic on almost every major highway on O’ahu has increased from 2012-2014. Moreover, the dearth of job creation on other parts of the island, namely the west side, has focused traffic into and out of downtown Honolulu, as shown in the first map.

This skewed traffic pattern, limited real estate, and inflexible road infrastructure has necessitated an above-ground railway system linking the west side of O'ahu with the burgeoning downtown area of Honolulu. This new organizing system seeks to rebalance the traffic system by reorganizing its resources, that is, by taking drivers and bus commuters off the road and onto the rail. O'ahu has only three major freeways, the H1, H2, and H3. The freeway H2 bottlenecks from the west into H1. Drivers and bus commuters are organized in such a way that peak hours of traffic are unavoidable. The new transit will conceivably provide an additional layer of organization to the currently

How much are the resources organized? There are 21 planned stations that run along the 20-mile span of track. The train stations are arranged to serve as many people as possible by concentrating them in the most densely populated areas.

Population Density

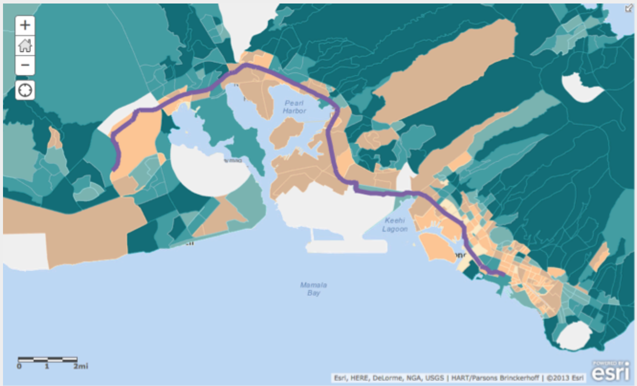


Population in Honolulu area is highly concentrated in lowland areas.

(Hawaii Dep. of Economic Development and Tourism. CC-BY-2.0.)

Darker areas represent high-density tracts while lighter areas represent low-density tracts. The densely-populated stretch from Keahi Lagoon to Honolulu Downtown, also has the highest density of traffic. It makes sense that this portion of the rail system constitutes almost half the number of total stops in just a quarter of the total mileage.

Honolulu Area Income per Household



Household income is lowest in the most densely populated areas.

(Hawaii Dep. of Economic Development and Tourism. CC-BY-2.0.)

Income per household also plays a vital role in how these stops were selected. The rail transit system predominately runs along areas of low-income neighborhoods (tan and brown indicates low income per household, while green indicates high income per household). This design principle embodies an assumption that people with lower incomes are more likely to rely on public transit.

When are the resources organized? As with any construction project of this magnitude, the organizing system was planned in detail before construction—down to the number of pillars, the amount of concrete, the imported steel for rail cars, etc. However, after construction excavation revealed ancient burial sites, the Native Hawaiian community demanded many changes to the project. The number of stops has remained the same but the route has changed dramatically.

How or by whom is it being organized? There are a number of interested parties with varying degrees of power. At the forefront, the government—that is, the State of Hawaii—makes the final decision. However, the people of Hawaii directly influence their decisions.

The protection of cultural resources, practices, and beliefs is important in Hawaii, both as a matter of law and of culture. Private archeology firms, state officials, and cultural descendants work together to reduce and mitigate impacts to archaeologically significant properties. The Oahu Island Burial Council, for instance, is a state council created to help protect *iwi kupuna* (ancestral bones). It stresses the importance of consulting recognized lineal descendants before any excavation for the rail project is carried out.

Where is it being organized? The “where” component of the organizing system is not as important for the scope of this analysis as other design questions. However, the physical nature of the project highly constrains how the system can be organized. The volcanic origin of O’ahu, does not allow for a below-ground rail system. The limited real estate, similarly, does not allow for a ground-level system. The sharp and steep volcanic ridges that cut across the island are barriers that limit where the rail system might go.

86. The Antikythera Mechanism

By Murray Maloney, 2 March 2014.

Overview. In 1900, a strange looking mechanical device was recovered from a shipwreck off of the island of Antikythera, Greece. Only in the 1970s was it determined that the device was an ancient mechanical computer that performed astronomical calculations; it had a manual crank control with a rate of one turn per day, forward or backward in time; its user interface presented calendrical, solar, lunar, and planetary positions.[1]

The Antikythera Mechanism

The Antikythera Mechanism exhibit at the National Archaeological Museum of Athens.

(Photo by Tilemahos Efthimiadis. CC-BY-2.0 license.)

The Antikythera Mechanism persists through time as a collection of artifacts and a model of intellectual achievement. Thought to have been constructed by Archimedes at Syracuse or by Posidonius at Rhodes, the mechanism was recovered from a ship wreck near the Greek island of Antikythera in 1900-1. The significance of the find only began to become apparent in the 1970s when researchers applied modern scanning technology.[2]

What is being organized? The Antikythera Mechanism was an arrangement of resource descriptions that represented a classical Alexandrian sol-lunar calendar, complete with an almanac of the positions of the sun, moon, known planets, and specific stars over time. The resource descriptions are represented simply by the measurements of the gears, and the corresponding information that is displayed on the front and rear panels, based on the position of those gears. These resource descriptions accounted for the range of known astronomical phenomena at the time.[3]

The organization of the mechanism consists of a main solar gear connected to a hand crank and a collection of gear trains that ultimately control the rotation of pointers indicating the calendar, lunar position and phases, the position of the sun and of all the known planets, and the nearest eclipse. The mechanism was housed in a wood frame box with bronze panels whose physicality was obviously intrinsic to the use of the device; the panels the back door was inscribed with what seems to be a user's guide.

The Antikythera Mechanism calculated the position of the moon by employing five gear trains to take into account the Saros, Metonic, Callipic, and Exeligmos cycles. Thus, it was able to predict the dates of solar and lunar eclipses.

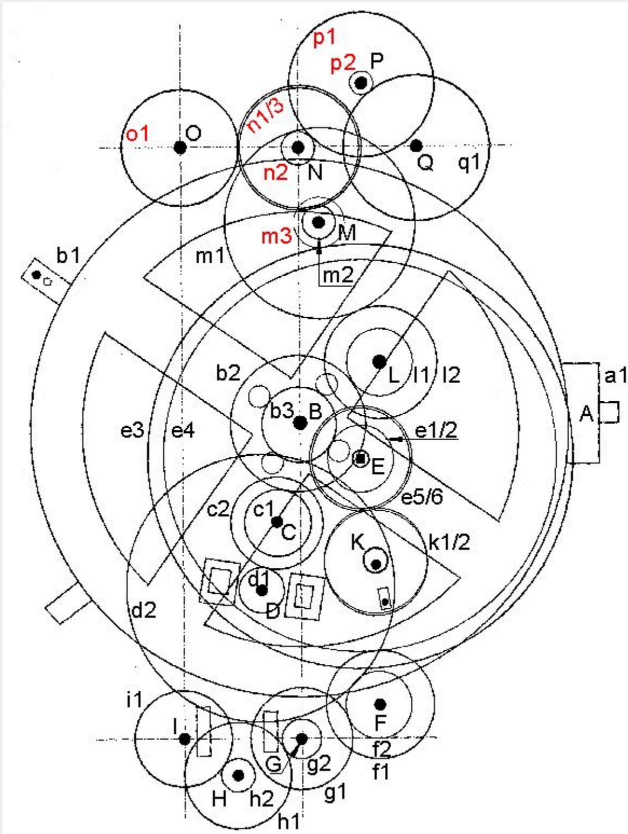
Today, the Antikythera is a collection of the eighty-two fragments that have been recovered from the ship wreck and sea bed, twenty three of which are evidently inscribed. The fragments have been dated to about 70 BCE based on the coincident presence of some coins from Pergamum and Ephesus that were recovered in the 1970s.

Why is it being organized? From a purely pragmatic perspective, the Antikythera Mechanism was a relatively portable computational device. that would have been used to accurately reckon a very specific calendar system, and to predict the cycles of days, months, years, and saro, as well as lunations, eclipses, and Olympic games. It would be an invaluable tool for astronomers, mathematicians, civil engineers, and cartographers of the time.

From a philosophical perspective, the Antikythera Mechanism was built to prove that it could be done. It represents a fulfillment of Aristotelian thought. Through the ages, the lure of scientific answers to the mathematical riddles presented among the patterns in the heavens has challenged our burgeoning intellects. The Antikythera Mechanism realized then-modern thinking on mathematics, engineering, astronomy and calendrical calculation in a portable mechanical computational device.[4]

How much is it being organized? Some of the major fragments are on display at the National Archaeological Museum of Athens; the others are stored.

The Gears



Gear arrangement.

(Wiki Commons.)

The Antikythera Mechanism is reported to have had about thirty gears within a frame whose size was less than the volume of a large book. The level of miniaturization and the precision of fabrication was not thereafter seen until the next millennium. The engineering and machining would have required trial models, accurate plans, and custom tooling. There have been various modern attempts to re-create the Antikythera Mechanism, or at least to re-create the model it seems to have manifested.[5]

When is it being organized? The person who operates the mechanism turns a hand-operated crank to establish a date, or contra-wise confirms the current date by taking sightings and comparing with the dial settings. The front face offers a solar-lunar calendar dial, a tropical zodiac dial, and an almanac dial with rising and setting times of various stars. The rear panel offers dials representing the five lunar cycles.

The Antikythera Mechanism

A recreation of the Antikythera Mechanism on display at the National Archaeological Museum of Athens.

(Photo by Tilemahos Efthimiadis. CC-BY-2.0 license.)

The organization of the engineering data required to build, operate, and maintain the Antikythera Mechanism is staggering to imagine, yet it pales in comparison to the organization required to collect and archive astronomical sightings on clay tablets for hundreds of years.[6] (See the sidebar, A Cuneiform Document at the Pergamon.)

The organization of the fragments of the Antikythera Mechanism is in the hands of the Bronze Collection of the National Archeological Museum in Athens.

How or by whom is it being organized? Ancient Chaldean, Greek, and Roman astronomers and engineers; modern divers, marine archaeologists, curators and researchers. In 1978, Jacques Cousteau led an expedition to the sea bed and returned some historical artifacts, that, while unrelated to the Antikythera Mechanism itself, provide additional historical context and may help date the discovery.

The Antikythera Mechanism Research Project is a collaboration of academic, industrial, and scientific researchers, who are applying some of the world's most advanced technology to study the capabilities and applications of the Antikythera Mechanism, as well as its historical context and significance.

Other considerations. From the perspective of one ship's unlucky captain and crew, the Antikythera Mechanism was likely just a piece of cargo, although it may have accompanied an equally unlucky passenger carrying the world's first computer to Caesar's court in Rome. It remains unknown how or why the device was aboard the ship or what fate befell it, but that is a story for researchers and historians to uncover in the fulness of time.

Notes: The following notes relate to this case study.

1. PBS aired Ancient Computer on April 3, 2013. The BBC aired Ancient Moon 'computer' revisited

2. The Antikythera Mechanism Research Project recently published The Inscriptions of the Antikythera Mechanism. 2016. Y. Bitsakis, M.G.Edmunds, A. Jones, *et alia* Almagest 7-1, May 2016

3. Cicero wrote about a similar device, created by Archimedes, in M. Tvlli Ciceronis de Republica Liber Primvs

   Gears from the Greeks. The Antikythera Mechanism: A Calendar Computer from ca. 80 B. C. Derek de Solla Price Transactions of the American Philosophical Society New Series, Vol. 64, No. 7 (1974), pp. 1-70

4. Aristotle's work on the subject *On the Heavens* (c 350 BCE) avers to the mathematical symmetry and perfection in the travels of the spheres, envisioning cycles and epicycles in motion.

   In 343 BCE, Aristotle was head of the Macedonian Academy, where he tutored Alexander and his future general, Soter Ptolemy. Following Alexander's conquest of Babylon in 331 BCE, he ordered Kallisthenes to organize the translation of all historical astronomical observations, initiating the transfer of the world's greatest collection of astronomical observations, dating back to 747 BCE. Within a year, Callippus had developed a new calendar, designating the summer solstice of 330 BCE as an epoch for astronomers and calendrical calculation. The Callipic cycle of 76 years less a day, equates to 27,759 days, and 940 lunations, is represented in the gearing of the mechanism.

   Ptolemy established his capital at Alexandria and founded a museum, spawning the need for a library, in the Platonic style. His successors, through to Cleopatra, added to the papyrus rolls. Mathematicians, astronomers, mechanical engineers, scientists; the most famous thinkers of the ancient world studied in the halls of the Library at Alexandria. Notable to us in this context are Euclid, Archimedes, Eratosthenes, Hipparchus, Aristarchus, and Posidonius.

   According to Pliny, the calendar reform of Julius Caesar, was assisted by Cleopatra's astronomer, Sosigenes, of Alexandria, who "brought the separate years back into conformity with the course of the sun."

5. In 2010, Andrew Carol built a Lego model of the Antikythera

Mechanism on a dare. John Pavlus wrote and directed a short film, Behind the Scenes: Lego Antikythera Mechanism.

Hublot, the Swiss maker of luxury time pieces, created a special edition Antikythera Watch. Hublot is also sponsoring ongoing research. See Return to Antikythera: A project of the Hellenic Ministry of Culture and Sports with support from the Woods Hole Oceanographic Institution

A simulation of the Antikythera Mechanism is available as an open source application on Github.

The Antikythera Mechanism Research Project maintains a list of Solid Models of the Antikythera Mechanism.

6. In his Almagest, Claudius Ptolemy marks the beginning of an epoch in recorded time, 1 Thoth 1 Nabonassar, with the coincident occurrence of a solar eclipse and the ascension of the Chaldean, King. Nabonassar in 747 BCE. (See the Almagest Ephemeris.) Nabonassar's calendar reform began a period of seven hundred years of meticulous record keeping, indexing, summarizing, and studying. The scientific study of astronomy based upon recorded observation is thought to have begun with Nabonassar. When we talk about the discipline of organizing, we can tip our hats to Nabonassar.

   John M. Steele (2000). Observations and Predictions of Eclipse Times by Early Astronomers. Kluwer Academic Publications. pp. 43–45.

   The British Museum stores the "Babylonian astronomical diaries," a highly systematized collection of ancient cuneiform texts that record periodic astronomical events, commodity prices and weather conditions over a period extending from 652 BCE to the 1st century BCE.

   Aaboe, Asger. The culture of Babylonia: Babylonian mathematics, astrology, and astronomy. The Assyrian and Babylonian Empires and other States of the Near East, from the Eighth to the Sixth Centuries B.C.E Eds. John Boardman, I. E. S.

Edwards, N. G. L. Hammond, E. Sollberger and C. B. F. Walker. Cambridge University Press, (1991)

Related Readings. See “Resources over Time”

87. Autonomous Cars

By Jason Danker, December 2015.

Overview. Automation in cars is nothing new. Automatic transmissions and cruise control have been around since 1939 and 1958 respectively, but these systems serve to aid, rather than replace, human drivers. What is new is a near future potential for fully autonomous cars, cars that are capable of full operation without an attending human driver.

While other vehicles, such as light rail and monorail trains, have been capable of fully automatic operation since 1967, these vehicles have the luxury of operating in closed environments and only need to be able to respond to a defined set of inputs. Autonomous cars do not have this luxury. In operating "in the wild," the systems guiding these cars may be forced to respond to any number of unanticipated situations. As the automation system cannot enumerate all possible situations, it must instead rely on continuous organization of its operating environment.

This is clearly a technical challenge, but it also raises ethical and legal issues. As autonomous cars act based on the organization of sensory inputs, the organizing systems are necessarily developed relative to ethical considerations, whether intentional or not. At the most basic level, the organizing system will direct the autonomous car in making decisions analogous to those posited in the trolley problem, a famous thought experiment in ethics that forces a choice between saving five endangered people or taking the life of an innocent person who had not been in danger. Beyond ethics, autonomous cars also raise legal questions: if an autonomous car crashes, who is liable for the damages?

What is being organized? An autonomous car will organize information about the car itself, the objects in its vicinity, and environmental conditions. The car must keep track of its movements, those of other objects, and the relative positions of itself and the other objects. It must organize this information within the environmental framework of lane markings, speed limits, road signs, traffic signals, weather and traffic conditions, and numerous other constraints. As autonomous cars become common, the cars will likely communicate with one another and this information will also need to be brought into the organizing system. The car will also need to organize, and likely prioritize, inputs from human occupants. Regardless of the exact implementation, the organizing system will necessarily limit what is worthy of organization: it is likely not possible, or desirable, to keep track of every insect in the vicinity of the car.

Why is it being organized? The car organizes its surroundings in order to safely navigate to a destination. While this is the primary interaction enabled by the organization, countless other interactions support this primary interaction. The supporting interactions fall into the two categories of prediction and reaction. The systems being developed by Google use the information that has been organized to predict what is most likely to happen next: "It predicts that the cyclist will ride by and the pedestrian will cross the street." The systems that have been launched by Tesla tend to be more reactionary: "Side Collision Warning further enhances Model S's active safety capabilities by sensing range and alerting drivers to objects, such as cars, that are too close to the side of Model S."

How much is it being organized? The extent of organization varies based on the implementation. While Google uses on-board sensors and extremely detailed street maps to implement self-driving functionality, Tesla's Autopilot relies on-board sensors and standard GPS data. While the exact extent of the organization is not publicly available information, Google has publicly stated "the system is engineered to work hardest to avoid vulnerable road users (think pedestrians and cyclists), then other vehicles on the road, and lastly avoid things that don't move." Given this, Google's categories, and their hierarchy, appear to be defined by their vulnerability.

When is it being organized? For information gathered by on-board sensors, organization takes place as objects enter and leave the vicinity of the autonomous car. The organization is ongoing as the car's surrounding and environment are constantly changing. In addition to the sensor data, autonomous cars also rely on map data which is organized in advance. Google's cars rely on specialized, highly detailed maps that are being developed as part of the self-driving car project and, as such, are unable to drive on roads that have not yet been mapped to the necessary level of detail. While Tesla's Autopilot also relies on maps, it uses standard GPS maps and is not similarly restricted.

How or by whom is it being organized? The car's computational processes are responsible for the organization. That said, the car is restricted to organizing within the organizing system implemented by the manufacturer. While Google and Tesla are two of the main companies in this space, many traditional automotive companies are also developing autonomous systems.

Where is it being organized? Except for map data, the organization takes place within the car's onboard systems. The organization must take place in the car itself due to the potential catastrophic consequences of a lag in information flow. Additionally, ensuring all organization takes place within the car provides greater security: a self-contained car is less susceptible to attack than a network dependent one.

Other considerations. While it is likely that fully autonomous cars will be technologically feasible within a few years, the cars may still require human interactions for legal reasons. This is clearly seen in Tesla's press release for Autopilot: "The driver is still responsible for, and ultimately in control of, the car." This human-in-the-loop design principle creates a legal buffer for autonomous car manufacturers by treating the "driver" as a "liability sponge" or "moral crumple zone." As articulated by Madeleine Elish and Tim Hwang, "the human in an autonomous system may become simply a component—accidentally or intentionally—that is intended to bear the brunt of the moral and legal penalties when the overall system fails."

While these issues will ultimately play out through a combination of court rulings and policy decisions, it is interesting to note that there is legal precedent that could either blame, or exonerate, the “driver” of an autonomous car. Drawing parallels to aviation automation, precedent suggests that the human “driver” will be held responsible for liability claims arising from the operation of the car. On the other hand, product liability law offers recourse for consumers when a company’s products fail. Many people have argued that this existing legal framework is sufficient to handle the liability issues brought up by autonomous vehicles.

Regardless of the legal complexities that will arise from specific incidents, autonomous cars have great potential to reduce car crashes and improve overall road safety. The promise of the autonomous technology, even for partially autonomous systems, is so great that the National Highway Traffic Safety Administration is proposing updates to its safety ratings that will penalize manufacturers that don’t include autonomous technologies in their vehicles.

88. IP Addressing in the Global Internet

By Andrew McConachie, December 2013.

Overview. Most people take for granted that the Internet just works. They connect their computer to the Internet, it gets an IP address, and they are able to communicate with a computer with a different IP address on the other side of the planet. How did their computer get the correct IP address? How does any computer or router get the correct IP address? How did the routers and other computers on the Internet get their IP addresses? Who decides which computers and which routers get which IP addresses?

What is being organized? At their simplest, an IPv4 address is a 32-bit series of 0′s and 1′s. They are resources that are born-digital, as they have no canonical physical representation. Their digital canonical representation, with which we have all become familiar, is called the "dotted quad" format and is 4 numbers between 0-255 separated with dots. For example, 169.229.216.200 is the IPv4 address for `www.berkeley.edu`.

Not all IP addresses are of equivalent classes. There are unicast, multicast, broadcast, and experimental IPv4 addresses, and unicast addresses can be either public or private. There are also two different versions of IP addresses currently in use on the Internet, IPv4 and IPv6. We will focus on IPv4 unicast public IP addresses, since these are not only the most common, but also the most important. This is roughly the range of IP addresses from 1.0.0.0 to 223.255.255.255, with some breaks in the middle for private IP address space.

Why is it being organized? IP addresses are the foundation of network connectivity and the Internet; they identify each device on a computer network and also serve as its address, so that routers and other devices can locate and communicate with it. You cannot get online without one. IP addresses can be represented into blocks, or subnetworks, using a prefix and a mask. For example, 169.229.216.0/24 represents all IP addresses in the range of 169.229.216.1 – 169.229.216.255. Internet routers do not have enough memory to hold routes for every individual IP address on the Internet. So by organizing the Internet into subnetworks based loosely on a hierarchical model, routers are able to determine the correct path for every destination in the network without actually storing every address in their memory. If the organization of IP addresses is not handled properly, Internet routers would exhaust their memory space and parts of the Internet would become unreachable.

How much is it being organized? Currently there is too much granularity in the global Internet routing table. For a router it takes the same amount of memory to store a subnetwork with 255 IP addresses as it does to store a subnetwork with 65536 addresses. So if our main concern is to minimize memory usage in Internet routers, thereby lowering operator costs and increasing stability, we want as little granularity as possible in the Internet routing table. The problem is that many organizations use non-contiguous IP subnetworks that cannot be aggregated into larger subnetworks. This results in routers having to store many small subnetworks instead of fewer larger subnetworks, which will eventually lead to memory exhaustion in older routers and possible reachability issues. Currently the full Internet routing table is approaching 500,000 routes. Most network engineers expect problems once the routing table grows past 512,000 entries, since router memory limitations are always at bit boundaries.

When is it being organized? IP addresses are organized once someone configures one on a device or sets up a Dynamic Host Configuration Protocol(DHCP) server. If an organization exhausts their supply of free IP addresses. it will have to make a request to the upstream provider or Regional Internet Registry(RIR) for more address space. In the early days of the Internet, large blocks of IP addresses were given to organizations, but this led to many of the addresses in these blocks not being used. We are now reaching a point where we no longer have new addresses to assign to organizations.

Markets are now emerging for organizations to buy and sell IP addresses, and the organizations who have held on to large amounts of unused addressing space stand to make significant revenue from selling their unused space. When these organizations sell their unused IP address space, they will break up large allocations into smaller subnetworks, thereby increasing granularity and further accelerating the growth of the Internet routing table.

How or by whom is it being organized? The Internet Corporation for Assigned Names and Numbers(ICANN) is currently responsible for initial allocation of IP addresses. They allocate 8 blocks of IP addresses to RIRs, who are then responsible for distributing allocations to organizations that request them. These organizations can then allocate IP addresses to smaller organizations, thus forming a loose hierarchy of organizations, where each level lower in the hierarchy receives a subset of the IP address space from the organization above it. ICANN no longer has any /8 blocks of IP addresses left to allocate to RIRs. Once all of the RIRs have exhausted their last allocations from ICANN, organizations will have to rely on secondary markets to increase their IP address space.

Other Considerations. The world of IP addressing will soon get a lot more interesting. The introduction of IPv6 as a replacement for IPv4 has been slow in coming and, while gathering momentum, continues at a snail's pace. As organizations start purchasing IP addresses from one another, we should expect increased granularity and decreased stability in the Internet routing infrastructure. Whether or not normal Internet users notice will ultimately be determined by how well equipment vendors and engineers expediently address the coming problems.

89. The Art Genome Project

By David Eicke, December 2014.

What is being organized? Artsy.net carries the ambitious mission of making "all the world's art" accessible to anyone with an Internet connection. This is not only challenging purely from a scale perspective, with the number of artworks in the world daunting even if it were not being incremented constantly, but it is also challenging in that "art" is a nebulous term. Creators of music and literature often refer to themselves and each other as "artists." The same goes for dancers and other performers. Will their works be included? The current collection seems to be mostly visual art, with some architecture and design objects included.

Artsy's mission is to be carried out by their Art Genome Project, which is the organizational engine that powers their search and interactions. The name was inspired by Pandora's project, as was their term for their organizing process: "genoming." Genoming is not yet automated and still costly, so Artsy selects the art that is to be "genomed" carefully. Their first priority is the works featured in galleries with whom Artsy has contracts. Galleries pay to have their work organized and searchable on the site. Those works, then, must be genomed quickly in order to keep the company running. Artsy's engine also takes in works from museums and other institutions who do not have contracts with them, but many of those institutions have image-rights concerns, and not all their artworks can be published. In other cases, the images of the works are simply too low-quality to be displayed.

Why is it being organized? Why organize art? The simplest answer is to educate. That said, art has been being organized into movements and -isms for a very long time. The Getty Foundation even created an authoritative art vocabulary called the Categories for the Description of Works of Art a few decades ago. At first glance, Artsy seems to be reinventing the wheel. However, the organizing system Artsy uses is unique in that it facilitates a special kind of interaction with its body of published works.

The way resources are organized on Artsy is a cross between a hierarchical structure and a graph structure. They have over 1,000 characteristics (which they call "genes") to describe their resources. These characteristics can have to do with art movements, formal qualities, techniques, subject, etc. The emphasis here, however, is on relationships between works of art. For example, one of the genes Artsy uses is "eye-contact," and if you have a photo taken last month where the subject is looking directly into the camera and an oil painting from hundreds of years ago where the subject's eyes are looking at the painter, those two can be one click away from each other. No other organizing system could facilitate that sort of easy link between two such disparate works.

This free-flowing linkage between works enables the "berry-picking" model of knowledge seeking, where a user searching for something doesn't necessarily have to know what he or she is searching for. A user could begin her exploration with only a vague notion that she enjoys this long-legged rhinoceros sculpture by Salvador Dali. She may not know what she likes about it, but she will see his other work there. Maybe she finds a painting she likes in the "other works by Dali" section, and she clicks on it. Then the characteristics of this painting are listed in the interface, and she is free to click on any one of them. She might click on "Surrealism" and find more works from that movement. She may click on "waterscapes" and find other oceanic imagery. She is free to explore and discover art in a self-directed way and free to discover what she likes and why she likes it. The director of Artsy's Art Genome Project says the system was intended to parallel a professor who is adept at "riffing" on things.

How much is it being organized? As mentioned above, Artsy currently uses over 1,000 characteristics ("genes") to describe its resources. These characteristics can describe anything from the art's form to the art's subject to the technique used to create the art. Experts assign these genes to the artworks and then assign those genes a weight from 0 to 100, depending on the salience of the characteristic within the work. Aside from the genes, the art is described in terms of physical dimensions (how much space it takes up), whether it has been sold or not, its gallery, its price (if for sale), its creation date, and, of course, who created it. Having such a rich set of descriptions has allowed Artsy to create a public API for developers to use all of this information as they see fit.

When is it being organized? Description of Artsy's resources is an ongoing process. Their ingested collection of art is much larger than their published collection. Most of the artworks are waiting to be genomed, with some of them waiting for permissions or image-rights paperwork to process. Another factor in determining when something is organized is the signing of new contracts with galleries. Works from galleries with contracts have first priority, and Artsy experts genome those works as they come in.

While these experts are assigning genes on a rolling basis, they are also drawing upon hundreds of years of art history scholarship when assigning them. For example, the Arsty experts did not come up with Dadaism as an organizational concept. So, in a way, some of these works were organized long ago.

How or by whom is it being organized? Artsy has a team of art historians and experts working to describe the resources that Artsy has ingested (and those that it will ingest). They have done some experiments with image-recognition software, but its descriptions are simply not rich enough to facilitate the sorts of interactions the organization is trying to facilitate. The strategy of employing experts has its obvious downsides, however. It does not scale well, and it is reminiscent of Yahoo's early strategy of employing librarians to describe web content. There will also be inevitable biases in human resource description.

Other considerations. With such a grand ambition, one thing that may stand in Artsy's way of becoming an authoritative organizing system in the art space is that they are for-profit. Even if they are able to avoid too much bias in the interest of revenue generation, the perception remains that they are less interested in classifying art for educational purposes and more interested in making money.

90. Making a Documentary Film

By Suhaib Syed, December 2013.

Overview. As part of a small crew, I was in pursuit of making a documentary film shedding light on the problems in the higher education system in India. We had traveled far and wide, capturing many thought-provoking stories, illuminating interviews, and shocking truths. Due to the relatively small crew and a tight schedule, we ended up with our raw footage being labeled in a generic format (MVI_1234 etc.). I, being the director, had the task of assisting the editor in renaming and reorganizing the files to make our lives easier, do justice to all the efforts that were put into capturing all the clips, and incorporate them in an impactful manner.

What is being organized? The primary resources being organized were the video clips (digital, shot on DSLRs) acquired during the shoot. In this context, they could be classified as passive resources having no real capability to produce any significant value on their own, and which had to be acted upon or interacted with to produce any effect. But the key problem here was to formulate usable resource descriptions based on the following resource properties:

Intrinsic static

> Date and time of creation, duration of the clip, type of external lighting used, camera used, lens used, exposure, ISO, white balance, frame rate, compression type

Extrinsic static

> Shot sequence number (assigned to each story element during story-boarding), shot movement type (dolly, follow focus, zoom, macro, etc.)

During this particular stage, the intrinsic and extrinsic dynamic properties did not play a large role in the resource descriptions.

We had done a lot of work on story-boarding and identified the right level of granularity so that we could capture each shot sequence separately, so we directly used the shot sequence number as an important part of the resource description. This helped us keeping our descriptions short and meaningful.

Additionally, we realized that the corresponding audio clips captured along with the video also had to be organized, but since the two were intricately linked to each other we decided to use the same name as the corresponding video clip, the only difference being the extension. We relied on the editing software to capture the intrinsic static properties of the audio files (e.g., bit rate and compression type).

Why is it being organized? Essentially, we were organizing these digital resources to find, identify, and select them so as to weave a powerful narrative enabling us to convey the truth in an impactful manner.

Hence, the interactions were directly with the primary resource.

The interactions that had to be supported by our organization scheme involved:

- Finding the clips related to a particular story-board section
- Selecting the best set of clips to be included in the film based on relevance to story, progression, continuation and several other inter-connected factors
- Manipulating the clip (i.e., color-correcting, white balancing, and stabilizing) to create an aesthetic effect
- Matching the video of a clip to corresponding audio recording

- Adding the right background score based on sentiment being portrayed in the clips and the progression of the story
- Providing subtitles in case of a foreign language or incoherent speech

How much is it being organized? Since the scope and size of our organizing system was relatively limited and all the resources were already available, we were able to make some bold decisions without causing a lot of problems. We formed a controlled, vertical vocabulary for resource description by deliberately choosing certain resource properties over others. Our main objective was to keep the description as short as possible and at the same time convey the most valuable information that would help us interact with the resources (i.e., the video clips).

We could have easily opted for a date- and time-stamp based id and every resource in a collection (i.e., clips specific to one camera) would have a unique identifier, but we realized that our cameras already attached this information to the file along with the technical details like frame rate, aperture, shutter speed, ISO, and white balance, which our operating system and editing software could easily capture, display, and search through, hence, we decided not to use these details.

We also decided not to include important lighting condition properties (kino-flo, LeikoLite, etc.) and location, because the first frame in most of our clips consisted of the clap-board which contained all of this information, and our editing software showed all the video files as thumbnails using first frame of the video.

Thus we leveraged all of these to form a controlled vocabulary that placed the shot sequence number first, followed by the take number followed by camera identifier (e.g., camA, camB, etc.). For instance: 2A_1_camB.

However, we did realize that these decisions were specific to our OS and video editing software and hence lacked interoperability.

When is it being organized? In our case, although we intended to organize the resources as soon as they were acquired, we failed and then came up with an organizing system after all the resources were acquired. We leveraged this fact to our benefit and formed a more specific description system.

How or by whom is it being organized? Ideally it is the role of the first assistant cinematographer (AC), even 2nd or 3rd AC (depending on the budget), to make sure all the file names are stored properly and all the cards properly backed up. But due to our limitations we (i.e., the director and cinematographer) collaborated to organize the set of raw footage.

Other considerations. One important consideration that we left out in the discussion was the need for certain people appearing in the documentary to have their identity hidden by means of facial blurring and voice modulation. Although we could not accommodate this interaction of identifying which clips had footage of people who did not want to reveal themselves, we could easily add the special effects over an entire sequence once all the clips were brought together.

91. The Dabbawalas of Mumbai

By Pratibha Rathore, December 2014.

Overview. The Mumbai dabbawala tiffin service is the source of much fascination from around the world, and I am no different: I worked in Mumbai for two years and used the services of dabbawalas to get my lunch box (called a "dabba") delivered from home to my office, which was about 44 miles away. Without the use of any technology or digital resources, this organizing system has been coordinating the delivery of home-cooked lunches to thousands of Indian office workers for over a century, charging just a small fee of $3-7 per month. The community of dabbawalas has been able to create value for its customers by optimizing and standardizing the principles of its operations and devising an organizing system that is down to earth and human-centric.

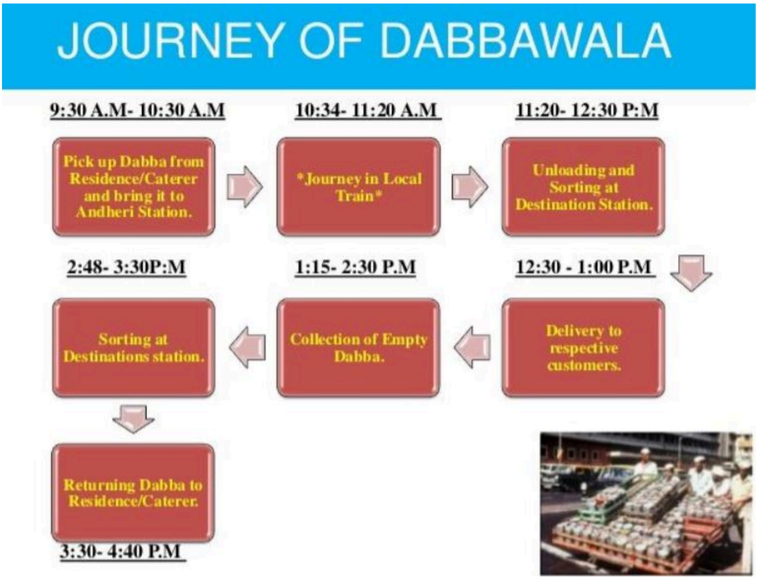
What is being organized? The primary resources in the dabbawala system are the dabbas that are delivered to respective customer's offices and organized using a simple but effective color-coding system. The secondary resource is the workforce, consisting of 5,000–6,000 people known as dabbawalas, who organize themselves and their supporting supply chain and logistics operations to deliver the dabbas to the right location and at the right time each day without failure. The dabbawala community, called the Mumbai Tiffin Box Suppliers Association (MTBSA), follows a flat organization structure, meaning the motivation to perform consistently is a matter of personal drive and accountability.

Why is it being organized? The primary reason people use the service of the dabbawalas is to eat a proper, home-prepared meal during lunch, a way to connect with their family while busy at work. The interactions supported by the dabbawala organizing system provide two significant benefits to the customers: managing their budgets while eating healthy, and leveraging time constraints. Most of the office-goers usually leave by 7 a.m. to commute from the suburbs of Mumbai, traveling south to the main commercial area of Mumbai and returning back home after 7 p.m. The railway network during the peak hours is jam-packed with commuters hanging onto the trains with one hand; therefore, carrying one's lunch at that time is not feasible. Most of the commuters cannot afford to eat takeout every day, and eating on the roadside is unhealthy and unhygienic. In addition, catering to the diverse food habits and taste needs of employees is very difficult for office canteens to manage. Thankfully, the dabbawala system solves all these problems with 100 percent customer satisfaction by delivering to each employee his lunch filled with food prepared at his home.

How much is it being organized? The Mumbai lunch box system is a successful and a socially sustainable enterprise. The number of dabbas delivered per day to offices and back home is around 300,000; that means 600,000 transactions per day. Although the number of transactions is very large, each person handles a small subset of transactions at a time. The scope of the organizing system and the scale of operations pretty much remain consistent, with the addition or deletion of few dabbas every month. Most interestingly, despite the lack of computers, mobile technology, or any automated processes, a dabba goes astray only once every two months, making less than one mistake in every 6 million deliveries. Now that's efficiency! The system is able to achieve consistency in its operations because of successful implementation of several organizing principles. Firstly, containers used to house the lunch boxes are of a standard shape and size. Second, the color coding done on the dabbas takes advantage of people's visual acuity, following a human-centric design approach. Third, the sequence of transactions to deliver each dabba from its source to destination and back to source is repeatable, predictable, systematic, and iterative in nature, enabling easy tracking and monitoring. Finally, governance within the community is achieved by instilling ethics, values, and principles in employees and by holding employees accountable at all times.

When is it being organized? The interactions between dabbawalas to deliver the dabbas follow a “hub and spoke” process model. During a dabba’s journey from kitchen to consumer, it is handled by between three and twelve different deliverymen. The typical day for a dabbawala begins at 9:30 a.m., and he spends about an hour collecting all the 25–30 dabbas from the assigned set of homes in his designated area. The households are expected to have the lunch box ready when he arrives for collection. When he is done with collection, he goes to the local train station and gathers with the other dabbawalas of his area. Next, the dabbas are sorted in the order of stops on that rail line and handed off to the dabbawala who is responsible for that particular station for delivery to their final destination. At every departure station, the dabbas are passed out according to their next destinations. The same process is repeated when returning empty dabbas back to homes.

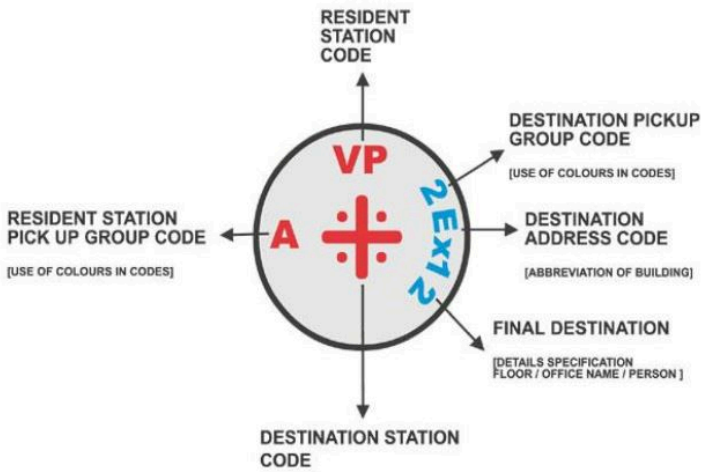
Dabbawalla Delivery Process



A model of the dabbawalla delivery process

How or by whom is it being organized? The key to this successful delivery management system is the color coding done on the dabbas. The dabbawalas use simple design measures such as signs, different colors, numbers, dashes, dots, letters, and simple symbols to indicate various parameters such as origination suburb, route to take, destination station, who is responsible, the street, building, floor, etc. As most of the dabbawals are illiterate, the choice of syntax for markings is done in such a way to ensure it is easy to understand and implement. The vocabulary used to implement and describe markings on the dabbas follows a standardized and self-descriptive process, thereby eliminating ambiguity and variability and making the organizing system more effective. Since only numbers and letters are used, the syntax for description of the primary resource (dabbas) is intentionally made to be independent of any local language, so that everyone can learn, understand, and process without any confusion, bias, or information overload.

Dabba Routing Codes



A breakdown of the coding system used to identify and route a dabba.

At each stage of the process, only one part of this code needs to be read, which works as a signal and thus allows picking up the right dabbas very quickly. It is also particularly efficient for traceability, since any dabbawala seeing a dabba knows which path it has to take. In case a dabba is lost or forbidden somewhere, any dabbawala is able to put it back on the right track. There is no need for the structure of color coding to be more granular than described above, as dabbawalas know the collection areas by heart. Furthermore, the process of adding a new resource to the organizing system is straightforward and structured. If a new resource—that is, a new customer—is added to the system, the dabbawala will do the complete journey to check the address of delivery and coordinate with other colleagues in the community to see who has a free place in his crate to add one more dabba. Once the sequence of delivery has been established and all the necessary stops for exchange decided, the address on the dabba is marked and it becomes part of the whole system.

Other considerations. It would be interesting to know if this delivery model could be used by other cities as the problem of longer commute and need for homemade food for lunch by office workers is always there in major cities. In my view, standardization of operations and understanding cultural and regional biases can provide opportunities for other cities to implement this model, at the same time providing jobs to many semi-skilled workforces.

92. Managing Information About Data Center Resources

By Hassan Jannah, December 2013.

Overview. Nowadays, there is an app for almost everything! Yet, we show little or no regard to what happens behind our shiny little screen until something breaks down and our lives descend to near chaos. That is the conundrum of IT guys. The truth is that IT solutions are, in many cases, fragile things that need constant care. This is no easy task. In fact, most of the cost and effort involved in IT solutions is maintenance. A million things could go wrong. Words like preventive maintenance, service monitoring, business continuity, and disaster recovery are examples of the different activities done to maximize availability and expedite troubleshooting. Everyone involved with these activities needs access to resources. Above all, they all need access to information.

What is being organized? IT data centers have both physical and digital resources. Physical resources include the facility (i.e., building), utilities, computer hardware (e.g., network switches, cables, servers, storage, etc.), and, also people. Digital resources are much fuzzier to define. A simplistic approach could classify them into data and applications. Each category can be further sub-classified into an entire ontology. The complexity increases when you consider the great number of potential resource types that can be created by combining physical and digital resources. Capturing, storing, and maintaining information about these resources is a big challenge. A lot of information can be retrieved from the resources themselves. Usually, each team responsible for supporting a certain group of resources would store information in spreadsheets and documents. More organized teams would use databases or knowledge management systems. More diligent organizations would have a central repository for everything.

What many fail to capture is the information about how all of these different clusters of resources are interconnected. That is often a much bigger and complex challenge. That information could be either buried deep in these systems (e.g., the user name used to run a certain service), or is stored in people's brains. The added value of an organizing system for data about data center resources can be multiplied if effectively organized information about their interactions.

Why is it being organized? Running an IT data center is complex, resource-intensive, and risky. Customers require around the clock availability of services with no room for failure. The consequences of such failures go beyond financial loss and customer dissatisfaction. They could affect people's safety and, even, national security. Cyber threats have become a constant threat for IT service providers, especially those that host highly sensitive data or serve critical operations. People can survive if their emails were inaccessible for an hour. However, what are the ramifications of a total failure of the IT infrastructure of the New York Stock Exchange? What if the airport systems of Heathrow airport failed? These are some of the conditions that IT data center managers must work in. Furthermore, technology advances have created highly diverse, complex, and integrated solutions. New resources are introduced frequently as old resources are retired. These activities require careful planning and execution to prevent the intricate ecosystem from crashing. Having all the information required to plan these activities would mitigate that risk.

Nevertheless, when something wrong does happen, having the required information is equally important to expedite fixing it. In fact, availability of information increases with the severity of the problem. How can you rebuild a system if you do not know how to connect its parts? How much are the resources organized? The granularity of the data required about data center resources varies between organizations and also between stakeholders of the same organization. The information can be classified into operational, and planning information.

How much is it being organized? Operational information is required for running day-to-day operations. These include information about resources and how they are interconnected. Many organizations put most of their focus on organizing operational information with high granularity. The granularity could be influenced by economic, political, and intellectual factors. Higher granularity means that more time and money are required to organize the information.

The level of granularity used to describe a resource type can be driven by the motives of the team leading the activity. For example, a hardware systems support team would invest more in building a robust organizing system for hardware systems and not focus on applications running on that hardware. Finally, the team's intellectual abilities and knowledge would influence the granularity of the system. As the boundaries between physical and digital resources fade, system designers could face some challenging questions. For example, servers are, traditionally, considered hardware resources. However, many organization have switched to virtual servers running on big machines. In such a case, how would you define a server? Is it the big machine or the individual virtual servers? Is it a physical resource or a digital resource? If you have a standby clone of a virtual server, would you consider both to be the same entity or not?

Planning information is usually required to make business decisions and is usually less granular. This could include information about the purchase and maintenance costs, contracts, hardware life-times …etc. Managers and planners could use this information to better plan for business activities, manage operational and capital costs, and make strategic decisions about the services and products the data center offers.

When is it being organized? Many data centers start building an organizing system of data about their resources based on existing resources. In such cases, building the system is the easy part. The real challenge is maintaining the information up-to-date in an ever-changing environment. Clear information life-cycle and change management processes are required in parallel with work processes to ensure information is updated.

How or by whom is it being organized? Based on the scope and level of granularity of the system, the number of resources could potentially be gargantuan. The organization must try to maximize the amount of information collected automatically using auto-discovery “agents” to keep updated information. Inevitably, other information, especially information describing interdependencies, will require human entry. The organization must have a clear and comprehensive governance framework that details the roles and responsibilities of different parties in adding, and maintaining information.

Other considerations. Most big companies in the past operated their own corporate data centers. Their organizing system might have a smaller scope. The emergence of global cloud service providers has extended the commoditization of IT products and services across the entire technology landscape; from the consumers all the way back to the servers that provide them. These providers will have a bigger scope due to the diversity and dynamic provisioning of their services.

93. Neuroscience Lab

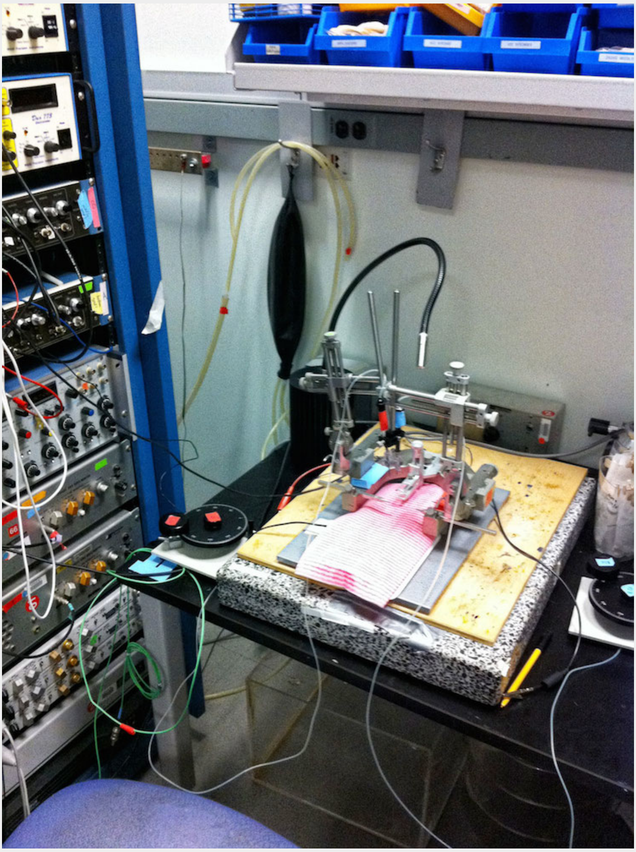
By Colin Gerber, December 2013.

Overview. A neuroscience lab is doing Parkinson's disease research in which they do experiments with rats. They use different types of rats, surgeries, and drugs for experiments and have to keep track of all this information for data analysis, publications, and lab inspectors.

The existing organizing system was developed before personal computers were prevalent and has slowly evolved over time. However, much of the underlying structure of the system still has its roots in pre-computer concepts. In order to update the system to incorporate more modern technologies what are the changes to the resources, their descriptions, and the systems structure that need to be made?

What is being organized? Resources in the current organizing system include rats, surgeries, experiments, drugs, and data recorded from the experiments. There are some other resources that could be incorporated into the organizing system.

Neuroscience Research Equipment



Physical resources in the author's lab are arranged to facilitate the precise accuracy of interactions required in medical research. In this photo, an array of amplifiers and filters for processing and recording rats' brainwave signals (left) is installed in a vertical rack that can be located close to the equipment used to perform surgeries.

(Photo by Colin Gerber. Used with permission.)

One such new resource is surgery techniques. Surgery techniques have historically been passed down by the master apprentice method and information was largely tacit knowledge that was held by the researchers performing the surgeries and not explicitly in the system. This was done because it is inherently difficult to store the intricacies of surgery in text and even more difficult for a new researcher to learn how to perform the surgery from textual information. The ability to store and annotate multimedia changes this however. It is now possible to make instructional videos for each type of surgery, add resource descriptions to the video file and store it in the organizing system.

There is also a resource that is treated as one resource through its entire lifetime when it may actually be two. When rats are originally brought into the organizing system they are treated as a manifestation of the rat resource type. Meaning the rats are interchangeable, you can use any rat from that group in your surgery. Once the surgery has been performed the rat is modified into a new resource instance. The specific rat the surgery was performed on now has a new set of resource descriptions.

Why is it being organized? Is the main purpose of the organization system to make sure the correct rats are used in each different experiment? Or is it to make sure the records are kept up to date for the lab inspectors? It could also be making data analysis and paper writing more efficient. These decisions will affect how many different types of resource descriptions are required and the granularity needed for those descriptions.

This system is just one of many organizing systems within a lab so deciding the scope and interactions it will have with the other organizing systems is very important. One important decision is if the system will support the training of new members of the lab or not. Having resources such as video recording of surgeries and experiments could enable teaching interactions for new researchers. But there are many other aspects of training a new researcher must go through, should these also be included in the organizing system? If so, it would make the system much more complex and expand the scope of the organizing system outside of surgeries and experiments but would keep all of the teaching resource in one system.

Another option would be to have a separate organizing system that is responsible for training material which is able to interact with the multimedia in the system that are relevant to training. This does not expand the scope of the system but would make the maintenance of it more difficult. Each time a surgery technique or experiment is changed two systems would have to be updated to take the changes into account.

How much is it being organized? The system is accessed by many types of users, each requiring a different type of interaction. The researchers need to search for the correct rat and surgery technique. The lab inspector needs to check for drug logs and make sure all the surgery methods and equipment are up to date. The principal investigator needs to see an overview of progress on projects.

Currently the system is organized in hierarchical categories where the top-level categories are surgery and experiments. This organization makes it easy to retrieve specific resources. However, the interactions normally performed with the system use resources from both sub-trees, which makes the hierarchical approach less than optimal.

A faceted classification approach could work well to enable these interactions. The facets would incorporate the original categories of surgery and experiments but also add facets for each common type of interaction. In this case different resource descriptions of the same resource will often be classified into different facets. These resource descriptions will often act as resources themselves. For example, a lab inspector is interested in retrieving the expiration date and times a drug was used in surgery, not the drug itself.

When is it being organized? In a neuroscience lab resource descriptions are often lost if they are not recorded at the time they are measured. For example, if a rat is weighed to calculate the correct dosage of a drug, both the dosage and the weight should be entered into the system. If the weight is not entered at the time of measurement it would be impossible to weigh the rat later and get the same result (as the rat changes weight over time.) This is a common problem, so as a rule all resources and descriptions should be entered into the system at the time they are acquired.

How or by whom is it being organized? The researchers working in the lab do all of the organizing. They are the ones creating new resources, descriptions and have the most knowledge about the resources and how they relate to each other.

Other considerations. Changing the system and entering all of the data at the time of measurement will initially cause more work for the researchers but will result in more accuracy for the interactions supported by the system and less retrieval work during data analysis and paper writing.

94. A Nonprofit Book Publisher

By Emily Paul, December 2014.

Overview. The New Press, a nonprofit book publisher with approximately 1,000 published titles, roughly 800 of which are actively in print and featured on the website, updated its book categories for use on thenewpress.com as part of a website redesign. Rather than fully adhering to an established book classification system, such as BISAC, which is commonly used in book retail, The New Press developed its own classification system. In addition to the standard goal of allowing readers to browse categories, this classification system is designed to represent the press's focus and mission. The New Press classification system employs a mix of principles and levels of granularity while incorporating some elements of the institutional categories from BISAC.

In order to gain some insight into how these dual goals affect usability, I ran user tests on a mockup of the website with the proposed categories. I conducted a think-aloud exercise in which the users verbalized their thoughts as they browsed through the categories and subcategories. I then asked the users to walk through where they would go for a particular book in response to a prompt from me that included the book's title, subtitle, and a brief description. Lastly, I asked the users about what their impressions were of The New Press after looking at the categories, whether they were confused by the categories, and which categories they would be interested in looking at if they visited the site.

What is being organized? The resource being organized is the digital presence of the books on thenewpress.com. The classification system is only used on The New Press website and is stored in a FileMaker database that pushes data to the website.

There is already a dedicated website classification system that this new system builds on. It is worth noting that the book records in the database also contain BISAC categories. These are entered so that they can be sent out to distribution and bookseller feeds that require the industry-standard categories. The BISAC categories are institutional categories created by the Book Industry Standards Group. The BISAC system is designed to reflect the interests and understanding of general readers. As such, the BISAC categories are informed by cultural categories and also influence cultural categories because of their broad adoption in the book industry. In addition to using some institutional categories from BISAC and mainstream cultural categories, The New Press is using cultural categories from specific groups, namely academics and political progressives, to connect with specific readers.

Why is it being organized? The books are being categorized to facilitate browsing by readers and supporters on The New Press website. In addition to the primary browsing interaction, the categories are also being used as an opportunity to position The New Press and to convey a sense of its mission.

How much is it being organized? For the purposes of The New Press website, books can be placed in multiple categories and subcategories, but all books will have at least one category designation. Because The New Press is not concerned with the physical presentation of the resources, the books can be placed in as many categories as are relevant. In contrast, library and bookstore classifications need to satisfy the uniqueness principle, because the book can only be located in one physical location.

Most of the categories are based on the subject matter of the books. A book's subject matter is an intrinsic static property because it does not change once it is published. However, the categories used to describe this subject matter may change over time as new categories are added to the classification system and retroactively assigned to previously published books. The book subject categories can generally be thought of as extrinsic and static because the threshold for changing them is higher than it is for more dynamic properties such as Current Season, Next Season, and Bestsellers. These categories are also included on the site in a separate section and are all extrinsic, dynamic properties because they are based either on time or sales, rather than intrinsic properties of the books.

The New Press classification system includes hierarchical categories, though only the subjects in which the press publishes more extensively have subcategories. In areas for which there are more books, the organization can be more granular without creating a subcategory that contains only one or a few books. Additionally, the greater institutional knowledge of the subject area enables the staff to make more specific distinctions within the broader subject category. One of the questions I explored in my user testing was whether these differentiations are necessary to support users' interactions with the books. If the users do not share the same level of knowledge in the subject it may not be useful, and may even diminish usability, to differentiate at the level of granularity provided by the subcategories.

Even at the top category level, there is a range of granularity and also a range of principles embodied in the categories. For example, History and Immigration are both top-level categories, but Immigration covers a more specific group of topics than History does. Most categories are based on the subject of the books, but there are several top-level categories based on other principles. These include Graphic Nonfiction, which refers to format; Primary Source Documents, which refers to the source material; and Biography, which refers to the genre of the book but does not express anything about its subject matter beyond the fact that it is about someone's life. Mixing category principles can be useful, particularly in a faceted system, which allows users to combine different categories to increase precision. In a faceted version of this system, a user could select Biography and Law in order to find biographies written about a judge or lawyer. Because books are assigned to all relevant categories in this system, this interaction is feasible at the logic level even though the current presentation does not allow it. If The New Press wanted to switch to a faceted presentation it would likely visually separate the categories into blocks based on the principles, so that users knew which facets they could pivot their searches on. This might include creating a genre section with Biography, Oral History, and Primary Source Documents as well as a geography section with the subcategories from World.

When is it being organized? Once the updated categories are finalized, all previously published books will be reviewed and assigned to new categories as necessary. Going forward, new books will be categorized on a seasonal basis and new categories may occasionally be assigned to previously published books on an ad hoc basis (this could be due to previous oversight in not assigning the category, or to the creation of a new category or subcategory). This system is flexible because books can be assigned to all relevant categories, so the introduction of a new category does not mean that all previous assignments will need to be changed. The subcategories also allow for flexibility because if one of these categories becomes more important over time, it can be changed at the presentation layer to a top-level category with minimal effort.

How or by whom is it being organized? The sales, marketing, and inventory manager assigns the categories, with input from the editorial and marketing teams. From time to time other departments, such as fundraising or publicity, may suggest a new category or category assignment for consideration. The categories are assigned in a FileMaker database in which the categories can be selected from a list of existing categories and subcategories. The category assignments in the FileMaker database are pushed to the website along with other book data.

Other considerations. Creating a classification system that can be widely understood is difficult to do. In this case, simplifying the system would support The New Press's goal of reaching a broad audience of readers. User testing revealed that the current category system may be hindering this because of issues with semantics, granularity, and structure. The structural issues are the most important to address because the inconsistent use of subcategories generated significant confusion during the user testing. By removing the subcategories and instead allowing expert users or those who know exactly what they are looking for to use search, the press could maximize the categories' relevance for general readers. This could be strengthened by an emphasis on using relevant keywords in the book descriptions that support searching. Despite some initial surprise from the test users about certain unusual top-level categories, I would argue that after simplifying other aspects of the system, the press could successfully keep some of these in order to represent its publishing areas and connect with like-minded readers. For example, Immigration and Criminal Justice are not top-level BISAC categories, but are easily understood by general readers and serve to highlight these important areas for The New Press. Biases in classification systems are unavoidable. While this can be negative, particularly when the organizers are not aware of the biases, it can also be harnessed positively and used to communicate a sense of the organization and its values. This needs to be approached thoughtfully and carefully and tested on users to understand how people outside the organization will interact with the system.

Afterword

ROBERT J. GLUSHKO

An Appendix to The Discipline of Organizing

We have reached the end of this book, but we are just at the beginning of “the discipline of organizing.” We hope that we have demonstrated why thinking of the art and science of organizing in a more abstract way can enable communication and cooperation across the numerous disciplines that are concerned with organizing, especially library and information science, computer science, informatics, law, economics, and business. Instead of just appropriating concepts and methods from these fields, we have tried to unify them, filling in the gaps between their complementary perspectives to yield a more comprehensive and generative understanding of how they fit together.

Now it is your turn. Perhaps you have a job in one of the fields we have brought together that involves organizing resources of one type or another. After reading this book, you surely will not approach that work the same way you did before. You will be able to apply the design patterns and principles of *The Discipline of Organizing* to make your existing organizing systems more capable, and will be able to create entirely new ones that fill the white space between the traditional categories.

We encourage people who read this book to contribute their own case studies at `DisciplineOfOrganizing.org`, and we expect to incorporate the most interesting and entertaining ones into this collection.

Dedication

ROBERT J. GLUSHKO

To Aristotle, Plato, Linnaeus, Condorcet, Wittgenstein…

Panizzi, Cutter, Raganathan, Svenonius…

Gibson, Norman, Rosch, Barsalou…

Adam Smith, Coase, Williamson…

Simon, Salton, Miller, Dumais…

Bush, Engelbart, Nelson, Berners-Lee…

…and the countless others whose diverse perspectives

we have synthesized in the discipline of organizing.

Acknowledgments

Acknowledgments

ROBERT J. GLUSHKO

Philosophers, scientists, designers, and many others have sought to make sense of how we organize our physical and intellectual worlds for over two thousand years. We owe a great general obligation to all of them, so we dedicated this book to them. However, it is more important to acknowledge more specifically the people who made *The Discipline of Organizing* happen. I think it is befitting of a book about organizing to be organized in making these acknowledgments, as follows in three categories:

The Motivators

Annalee Saxenian, the Dean of the UC Berkeley School of Information, challenged me in 2005 to teach the "Information Organization and Retrieval" course required of all entering graduate students and provided me with a supportive environment in which to do it. The lecture notes of my predecessors, Berkeley colleagues Marti Hearst, Ray Larson, and Mark Davis, provided important intellectual scaffolding as I developed my own syllabus and lectures.

When I discovered the little red book by Elaine Svenonius, *The Intellectual Foundation of Information Organization*, my mind opened up to library and information science. I aspired to write a book that could build on and broaden those foundations to connect with my own background in cognitive and computer science. A few months later when I met Elaine I was very pleased when she endorsed this ambitious effort.

I have been continually encouraged by faculty members and deans whenever I talked about this project at Schools of Information or similar academic units. These include the U.S. universities of Indiana, Kentucky, Michigan, and North Carolina, Canadian universities of Toronto and Western Ontario, and European universities in Vienna and Berlin. In particular, I would like to thank Colin Allen, Ron Day, Miles Efron, Thomas Finholt, Dan O'Hair, Margaret Hedstrom, Michael Jones, John King, Kathryn LaBarre, Kelly Lyons, Gary Marchionini, Jerry McDonough, Allen Renear, Seamus Ross, Victoria Rubin, Michael Seadle, and Linda Smith. I especially appreciate the encouragement that Deans Marchionini and Seadle gave to Ryan Shaw and Vivien Petras, two of the principal authors of this book. I apologize to those of you that I have forgotten to list here.

Margy Avery of The MIT Press has pushed hard when she needed to and has been very receptive when I needed her to be.

The Contributors

It took me four years of teaching the IO & IR course at Berkeley before I knew enough (or too little) to think I could put together a book that might replace the diverse set of textbooks that course was using. I did not realize at the time how much I was learning from these teaching assistants, and I thank them for not making it obvious to me. Later on, after the book project was underway, my teaching assistants were invaluable in pointing out problems with the book, often proposing their solutions as well.

Almost exactly three years ago the project to write this book began in a graduate seminar whose goals were to define the topical coverage and structure of the book, and then to write chapters starting with my course lecture notes. Among the courageous students in that seminar were many authors of the book being published here: Rachelle Annechino, J.J.M. Ekaterin, Ryan Greenberg, Jess Hemerly, Michael Manoochehri, Sean Marimpietri, Kimra McPherson, Karen Nomorosa, Hyunwoo Park, Dan Turner, and Longhao Wang. Nick Doty, Mohit Gupta, Erin Knight, and Joyce Tsai also contributed during this start-up period.

In Spring 2011 Erik Wilde and I conducted a seminar titled "Principles and Patterns of Organizing Systems" to refine the key concepts of the evolving book. This seminar added Brendan Curran, Krishna Janakiraman, Julian Limon, Rowyn McDonald, Elisa Oreglia, Monica Rosenberg, Karen Rustad, Bailey Smith, Leslie Tom, and Anne Wootton to the growing set of student contributors. Leslie gets credit for the book's title.

In Spring 2012 Andrea Angquist, Jacob Portnoff, and Brian Rea, supervised capably by Anne Wootton, were essential editorial assistants in my end-to-end effort to rewrite the drafts of Chapters 1-7 to improve their conceptual integration and continuity.

I used draft chapters of the book in my IO & IR course three times, beginning in Fall 2010. The final version of the book in 2013 barely resembles those early drafts, which means that many students suffered to improve the book. But they did not suffer passively. Many students submitted problems with the Twitter hashtag #tdofix, and submitted examples using #tdoexample, which benefited the book greatly but which surely confused their regular Twitter followers.

Many other Berkeley students did important work on the book. Jen Wang designed the cover; Divya Anand, Ajeeta Dhole, Christina Pham, and Raymon Sutedjo-The did the illustrations; Lisa Jervis, Shohei Narron, and Anne Wootton worked on the extensive bibliography. A group of students whose work does not appear in the printed book but whose efforts will be revealed in future ebooks include Luis Aguilar, Fred Chasen, Philip Foeckler, Jake Hartnell, Eliot Nahman, and AJ Renold.

Eliot Kimber showed me that it was possible to write a book that could be published simultaneously in print and in ebooks. It has not turned out to be as simple as someone as talented as Kimber can make it seem, but I am grateful to Eliot for convincing me that I should try to do it. With help from Bob Stayton, Adam Witwer (and O'Reilly Media) we will get there.

I must also thank Christine Borgman of UCLA for bringing a group of energetic and thoughtful UCLA graduate students into the project. Two of them, Matt Mayernik and Alberto Pepe, are contributing authors. Amelia Acker, Jillian Wallis, and Laura Wynholds taught me a great deal about libraries and archives, and I am certain they tried to teach me much more than I was able to learn.

Many people read draft chapters and were thankfully unsparing in their criticism because they wanted to make this book as good as it could be. Thank you Scott Abel, Larry Barsalou, Marcia Bates, Christine Borgman, Michael Cohen, David Kirsh, Jeff Elman, Rob Goldstone, Jonathan Grudin, Ben Hill, Mano Marks, Patrick Schmitz, Elaine Svenonius, Jeff Zych, and everyone else whom I have carelessly forgotten.

Few books have been as battle tested before they went to print as this one. Let me thank those who have been willing to teach from a book that did not entirely exist: Jane Greenberg, Irith Hartman, Lauren Plews, Sarah Ramdeen, Christian Sandvig, Emily Seitz, Isabelle Sperano, Konstantin Tovstiadi, Hong Zhang, and especially Vivien Petras and Ryan Shaw who went to battle with (and for) this book multiple times.

The Essentials

The third and final category of acknowledgments is for people who were essential, without whom this project would never have been finished.

Jess Hemerly and Kimra McPherson joined the project in the “first campaign” of Spring 2010, worked tirelessly through that summer to make chapter drafts course-worthy, and served as teaching assistants in Fall 2010 when the book was first tested with students. They helped me believe that there might be a book in there somewhere when it took a lot of faith to see that.

Erik Wilde taught me much through our multi-year collaboration and dialectic when he was on the Berkeley faculty from 2006-2011. Erik made me understand the elegance and great scope of the word “resource,” which became the central concept in this book. His meticulously annotated reviews of many chapters from a computer science perspective helped inspire the idea of discipline-tagged endnotes.

Ryan Shaw and Vivien Petras, both young professors at schools a long distance from Berkeley, found courage in themselves and had confidence in the draft book coming out of Berkeley in 2011—first to teach with it, and then to help write it, becoming the primary reviewers of my chapters and the first authors of Chapters 8 and 9.

Murray Maloney joined the project in April 2012 as copy editor, but we together soon recognized that his nearly three decades of SGML, XML, and publishing experience were too valuable not to exploit further for the benefit of this book. Without Murray’s work as the markup and production editor, indexer and glossary-maker, there would be too much work left to do and no one capable of doing it as well as Murray has. Somehow along the way he also found time to make important intellectual contributions as a co-author in Chapters 5 and 8.

Finally, I want to thank Pam Samuelson. She has been far too patient with me as I talked with her, to her, and at her for three years while this book was being written, who turned many quarter-baked ideas into half-baked ones, and who turned many half-baked ones into cornerstones of this book. Most importantly, she has helped me focus on this book and get it finished when it would have been easy to give up on it. I promise not to take on another book project anytime soon because Pam has suffered enough for this one.

Robert J. Glushko, 31 December 2012

Credits and Notices

ROBERT J. GLUSHKO AND MURRAY MALONEY

Editor: Robert J. Glushko

Principal Authors: Robert J. Glushko, Jess Hemerly, Murray Maloney, Kimra McPherson, Robyn Perry, Vivien Petras, Ryan Shaw, and Erik Wilde

Contributing Authors: Rachelle Annechino, Matt Earp, J.J.M. Ekaterin, Graham Freeman, Ryan Greenberg, Daniel Griffin, Carl Lagoze, Ian MacFarland, Michael Manoochehri, Sean Marimpietri, Matthew Mayernik, Karen Joy Nomorosa, Hyunwoo Park, Alberto Pepe, Jordan Shedlock, Isabelle Sperano, Daniel D. Turner, and Longhao Wang

Case Study Authors: Daniel Brenners, Gracen Brilmyer, Jason Danker, David Eicke, Colin Gerber, Shaun Giudici, Emilie Hardman, Hassan Jannah, Sandra Lee, Carlo Liquido, Ian MacFarland, Andrew McConachie, Emily Paul, Pratibha Rathore, Becca Stanger, and Suhaib Syed

Bibliography Editors: Lisa Jervis, Shohei Narron, and Anne Wootton

Cover Designer: Jen Wang

Illustrators: Divya Anand, Ajeeta Dhole, Robyn Perry, Christina Pham, and Raymon Sutedjo-The

CSS Page Designs: Nellie McKesson, Adam Witwer, Murray Maloney

User Interface Consultant: Isabelle Sperano

DocBook Consultants: Bob Stayton and Jirka Kosek

Markup & Production Editor: Murray Maloney

Notices

Published by (2020): Robert J. Glushko. San Francisco, CA.

Published by (2014-2016): O'ReillyMedia. Sebastopol, CA.

Published by (2013): The MIT Press. Cambridge, MA.

Produced in the United States of America.

Library of Congress Cataloging-in-Publication Data.

The Discipline of Organizing (4th ed.)/ edited by Robert J. Glushko.
Includes bibliographical references and index.

1. Information organization.
2. Information resources management.
3. Metadata.

I. Glushko, Robert J., editor of compilation.

Z666.5.D57 2013
025—dc23

10 9 8 7 6 5 4 3 2 1 0

Dedication

Dedication

MURRAY MALONEY

Dedication

Aesop, Homer, Aristotle, Alexander, Ptolemy Soter, Aristophanes…

Hieronymus, Gutenberg, Pfister, Petrucci, Manutius, Braille…

Bush, Engelbart, van Dam, Nelson, Hart…

Saltzer, McIlroy, Osanna, Kernighan, Knuth…

Allen, Bosak, Fogel, Maler, Walsh…

Adler, Clark, DeRose, Goldfarb, Kimber, Newcombe, Orchard, Wohler…

Berners-Lee, Connolly, Raggett, Pemberton, Berjon…

Lilley, Garrish, Kerscher, Miner, Sargent, Zilles…

…and the countless others whose diverse perspectives

continue to shape… the written word.

Bibliography

MURRAY MALONEY

An Appendix to The Discipline of Organizing

Note

Entries begin with an identifier based upon principal author and date of publication; in cases where a principal author is associated with multiple works in a given year, a suffix is appended to differentiate. These identifiers serve as hypertext links to guide the gentle reader back to a corresponding primary citation within the chapters of the book.—MM

A

[Aalbersberg2011] Ijsbrand Jan Aalbersberg. 2011. *"Supporting Science through the Interoperability of Data and Articles".* *D-Lib Magazine*. Vol. 17. no. 1/2.

http://www.dlib.org/dlib/january11/aalbersberg/01aalbersberg.html.

[Abel2014] Scott Abel and Rahel Anne Baillie. 2014. *The Language of Content Strategy.* XML PRESS.

[Ackoff1989] Russell L. Ackoff. 1989. *"From data to wisdom."*. *Journal of applied systems analysis.* 16. no. 1. pp. 3-9.

[Agrawal1989] Rakesh Agrawal, Alexander Borgida, and H. V. Jagadish. 1989. *"Efficient Management of Transitive Relationships in Large Data and Knowledge Bases"*. *SIGMOD '89: Proceedings of the 1989 ACM SIGMOD international conference on management of data.* pp. 253-262.

[Allmendinger2005] Glen Allmendinger and Ralph Lombreglia. 2005. *"Four Strategies for the Age of Smart Services"*. *Harvard Business Review.* October. http://hbr.org/2005/10/four-strategies-for-the-age-of-smart-services/ar/1.

[Anderson2008] Chris Anderson. 2008. *"The end of theory: The data deluge makes the scientific method obsolete"*. *Wired.*

http://www.wired.com/2008/06/pb-theory/

.

[Anderson2001a] James D. Anderson and José Pérez-Carballo. 2001. *“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 and Management*. Vol. 37. pp. 231-254. http://www.sciencedirect.com/science/article/pii/S0306457300000261. doi: 10.1016/S0306-4573(00)00026-1.

[Anderson2001b] James D. Anderson and Jose Perez-Carballo. 2001. *“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 and Management*. Vol. 37. pp. 255-277. http://www.sciencedirect.com/science/article/pii/S0306457300000467. doi: 10.1016/S0306-4573(00)00046-7.

[Anderson2001c] Stephen R. Anderson. 2001.

"Morphololology". In The MIT Encyclopedia of the Cognitive Sciences. Robert Wilson and Frank Keil, (Eds.). 562-563. Cambridge, MA.

A BRADFORD BOOK.

[Anderson2003] John Anderson. 2003. *Art Held Hostage: The Battle over the Barnes Collection.* New York.

W. W. NORTON & COMPANY.

[Anderson2006] James Anderson and Melissa Hoffman. 2006. *"A Fully Faceted Syntax for Library of Congress Subject Headings". Cataloging & Classification Quarterly.* Vol. 43. no. 1. pp. 7-38.

[Apte1995] Uday M. Apte and Richard O. Mason. 1995. *"Global Disaggregation of Information-Intensive Services". Management Science.* Vol. 41. no. 7. pp. 1250-1262. http://www.jstor.org/stable/2632780.

[Arasu2001] Arvind Arasu, Junghoo Cho, Hector Garcia-Molina, Andreas Paepcke, and Sriram Raghavan. 2001. *"Searching the Web". ACM Transactions on Internet Technology.* Vol. 1. no. 1. pp. 2-43. doi: 10.1145/383034.383035.

[Argott2009] *The Art of the Steal*. 2009. Don Argott.

NEW YORK: IFC FILMS.

[Aristotle350BC] Aristotle. 350 BC. *On the heavens. (De Caelo et Mundo)*. Ancient Greece. https://archive.org/details/decaeloleofric00arisuoft.

[Arthur1992] Paul Arthur and Romedi Passini. 1992. *Wayfinding: People, Signs and Architecture*. New York.

MCGRAW-HILL.

[Atran1987] Scott Atran. 1987. *"Ordinary Constraints on the Semantics of Living Kinds: A Commonsense Alternative to Recent Treatments of Natural-Object Terms"*. *Mind & Language*. Vol. 2. no. 1. pp. 27-63. http://onlinelibrary.wiley.com/doi/10.1111/j.1468-0017.1987.tb00107.x/abstract. doi: 10.1111/j.1468-0017.1987.tb00107.x.

[ATT2011] . 2011. *"An AccuWeather Cloudlet Answers a Hail of Data Requests"*. http://www.business.att.com/content/customertestimonial/Case_Study-_AccuWeather_4.7.11.pdf.

[Atzori2010] Luigi Atzori, Antonio Iera, and Giacomo Morabito. 2010.

“The Internet of Things: A survey”. Computer Networks. Vol. 54. pp. 2787-2805. http://www.sciencedirect.com/science/article/pii/S1389128610001568. doi: 10.1016/j.comnet.2010.05.010.

[Aufderheide2011] Patricia Aufderheide and Peter Jaszi. 2011. *Reclaiming Fair Use: How to Put Balance Back in Copyright.* Chicago.

UNIVERSITY OF CHICAGO PRESS.

B

[Baeza-Yates2011] Ricardo Baeza-Yates and Berthier Ribeiro-Neto. 2011. *Modern Information Retrieval: The Concepts and Technology behind Search.* Boston, MA.

ADDISON WESLEY.

[Bailey2007] Charles W. Bailey. 2007. *“Open Access and Libraries”. Collection Management.* Vol. 32. no. 3/4. pp. 351-383.

[Baker1962] Keith M. Baker. 1962. *“An unpublished essay of Condorcet on technical methods of classification”*. *Annals of Science*. Vol. 18. no. 2. pp. 99-123.

[Banzhaf2009] Wolfgang Banzhaf. 2009. *“Self-organizing Systems”*. *Proceedings of Encyclopedia of Complexity and Systems Science*. pp. 8040-8050.

[BarIlan2008] Judit Bar-Ilan. 2008. *“Informetrics at the beginning of the 21st century—A review”*. *Journal of Informetrics*. Vol. 2. no. 1. pp. 1-52. http://www.sciencedirect.com/science/article/pii/S1751157707000740. doi: 10.1016/j.joi.2007.11.001.

[Barsalou1983] Lawrence W. Barsalou. 1983. *“Ad hoc categories”*. *Memory & Cognition*. Vol. 11. no. 3. pp. 211-227. http://www.ncbi.nlm.nih.gov/pubmed/6621337.

[Barta2009] Patrick Barta. 2009. *“Shifting the Right of Way to the Left Leaves Some Samoans Feeling Wronged”*. *The Wall Street Journal*. 2009. http://online.wsj.com/article/SB125086852452149513.html.

[Bartok1981] Béla Bartók and Albert B. Lord. 1981. *The*

Hungarian Folk Song.. Edited by Benjamin Suchoff. Translated by M.D. Calvocoressi. Annotated by Zoltán Kodály..

STATE UNIVERSITY OF NEW YORK PRESS.

[Bates2005] Marcia J. Bates. 2005. *"Information and knowledge: an evolutionary framework for information science". Information Research.* Vol. 10. no. 4.

[Batley2005] Sue Batley. 2005. *Classification in Theory and Practice.* Cambridge, UK.

CHANDOS PUBLISHING.

[Batt2000] Rosemary Batt. 2000. *"Strategic Segmentation in Frontline Services: Matching Customers, Employees, and Human Resource Systems". International Journal of Human Resource Management.* Vol. 11. no. 3. pp. 540-561. doi: 10.1080/095851900339756.

[Batten1951] W. E. Batten. 1951. *"Specialized Files for Patent Searching". In Punched Cards: Their*

Applications to Science and Industry. Robert S. Casey and James W. Perry, (Eds.). 169-181. New York.

REINHOLD PUBLISHING CORPORATION.

[Battistella1996] Edwin Battistella. 1996. *The Logic of Markedness*. Oxford.

OXFORD UNIVERSITY PRESS.

[Battles2003] Matthew Battles. 2003. *Library: An Unquiet History*. New York.

W. W. NORTON & COMPANY.

[Bean2001] Carol A. Bean and Rebecca Green. 2001. *Relationships in the Organization of Knowledge*. Norwell, MA.

KLUWER.

[Bell1970] Barbara Bell. 1970. *"The oldest records of the Nile floods."*. *The Geographical Journal*. 136. no. 4. pp. 569-573. .

[Bentivogli2000] Luisa Bentivogli and Emanuele Pianta. 2000. *"Looking for lexical gaps"*. *Proceedings of Euralex-2000 International Congress*.

[Bergmark2002] Donna Bergmark, Carl Lagoze, and Alex Sbityakov. 2002. *"Focused Crawls, Tunneling, and Digital Libraries"*. *Proceedings of the 6th*

European Conference on Research and Advanced Technology for Digital Libraries. pp. 91-106.

[Berlin2014] Brent Berlin. 2014. *Ethnobiological classification: Principles of categorization of plants and animals in traditional societies..*

PRINCETON UNIVERSITY PRESS.

[Berners-Lee1998] Tim Berners-Lee. 1998. *"Cool URIs don't change".* *World Wide Web Consortium (W3C).* http://www.w3.org/Provider/Style/URI.html.

[Berners-Lee2001] Tim Berners-Lee, James Hendler, and Ora Lassila. 2001. *"The Semantic Web".* *Scientific American.* May.

[Biasiotti2008] Mariangela Biasiotti, Enrico Francesconi, Monica Palmirani, Giovanni Sartor, and Fabio Vitali. 2008. *"Legal informatics and management of legislative documents".*

GLOBAL CENTRE FOR ICT IN PARLIAMENT.

[Bitner1992] Mary Jo Bitner. 1992. *"Servicescapes: The impact of physical surroundings on* *customers and employees". Journal of Marketing*. Vol. 56. no. 2. pp. 57-71.

[Bizer2009a] Christian Bizer. 2009. *"The Emerging Web of Linked Data". IEEE Intelligent Systems*. Vol. 24. no. 5. pp. 87-92. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5286174.

[Bizer2009b] Christian Bizer, Tom Heath, and Tim Berners-Lee. 2009. *"Linked Data—The Story So Far". International Journal on Semantic Web and Information Systems*. Vol. 5. no. 3. pp. 1-22. doi: 0.4018/jswis.2009081901.

[Blanchette2002] Jean-François Blanchette and Deborah G. Johnson. 2002. *"Data Retention and the Panoptic Society: The Social Benefits of Forgetfulness". The Information Society*. Vol. 18. Iss. 1.

[Blanzieri2009] Enrico Blanzieri and Anton Bryl. 2009. *"A survey of learning-based techniques of email spam filtering". Artificial Intelligence Review*. Vol. 29. no. 1. pp. 63-92. doi: 10.1007/s10462-009-9109-6.

[Blei2012] D. M. Blei. 2012. *"Probabilistic topic models". Communications of the ACM*. 55. 4. 77-84.

[Board2002] , , and . 2002. *Effectiveness and Impact of Corporate Average Fuel Economy (CAFE) Standards.* Washington, DC.

THE NATIONAL ACADEMIES PRESS.

http://www.nap.edu/openbook.php?isbn=0309076013.

[Bolshakov2004] Igor A. Bolshakov and Alexander Gelbukh. 2004. *"Synonymous Paraphrasing Using WordNet and Internet". Proceedings of NLDB: International Conference on Applications of Natural Language to Information Systems.* pp. 312-323.

[Borges1952] Jorge Luis Borges. 1952. *"The Analytical Language of John Wilkins" (El idioma analitico de John Wilkins). In Otras Inquisiciones (1937–1952).* Buenos Aires, Argentina.

[Borgman2000] Christine L. Borgman. 2000. *From Gutenberg to the Global Information Infrastructure: Access to Information in the Networked World.* Cambridge, MA.

THE MIT PRESS.

[Borgman2011] Christine L. Borgman. 2011. *"The Conundrum of Sharing Research Data". Journal of the American Society for Information Science and Technology.* pp. 1-40. http://papers.ssrn.com/abstract=1869155.

[Boroditsky2003] Lera Boroditsky. 2003. *"Linguistic Relativity". In Encyclopedia of Cognitive Science.* Hoboken, NJ.

WILEY.

[Boroditsky2010] Lera Boroditsky. 2010. *"Lost in Translation". The Wall Street Journal.* 2010-07-23.

http://online.wsj.com/article/SB10001424052748703467304575383131592767868.html.

[Boroditsky2011] Lera Boroditsky. 2011. *"How Language Shapes Thought". Scientific American.* February. http://www.sciamdigital.com/index.cfm?fa=Products.ViewIssuePreview&ARTICLEID_CHAR=94C85092-237D-9F22-E874366AD6B49809.

[Bower2001] Jim Bower and Andrew Roberts. 2001. *"Developments in International Museum and Cultural Heritage Information Standards".* Paris.

INTERNATIONAL COMMITTEE FOR DOCUMENTATION (CIDOC) OF THE INTERNATIONAL COUNCIL OF MUSEUMS (ICOM).

[Bowker2000] Geoffrey C. Bowker and Susan Leigh Star. 2000. *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA. THE MIT PRESS.

[Brailsford1999] David F. Brailsford. 1999. *“Separable Hyperstructure and Delayed Link Binding”*. *ACM Computing Surveys*. Vol. 31. (4es).

[Bray2005] Tim Bray. 2005. *“On Language Creation”*. *XML 2005*.

[Briet1951] S. Briet. 1951. *“Qu’est que la documentation?”*. *Editions Documentaires Industrielles et Techniques*. Paris.

[Brin2009] Sergey Brin. 2009. *“A Library to Last Forever”*. *The New York Times*. 2009-10-09. http://www.nytimes.com/2009/10/09/opinion/09brin.html.

[Brown2002] John Seely Brown and Paul Duguid. 2002. *The Social Life of Information*. Cambridge, MA. HARVARD BUSINESS PRESS.

[Brown2009] Bruce C. Brown. 2009. *How to Stop*

E-Mail Spam, Spyware, Malware, Computer Viruses, and Hackers from Ruining Your Computer or Network: The Complete Guide for Your Home and Work. Ocala, FL.

ATLANTIC PUBLISHING GROUP INC.

[Brown2010] Dan Brown. 2010. *Communicating design: developing web site documentation for design and planning.*

NEW RIDERS.

[Bruner1957] Jerome S. Bruner. 1957. *"Going beyond the information given".* *In Contemporary approaches to cognition.* J.S. Bruner, Brunswik E, L. Festinger, F. Heider, K.F. Muenzinger, C.E. Osgood, and D. Rapaport, (Eds.). 41-69.

HARVARD UNIVERSITY PRESS.

http://www.jimdavies.org/summaries/bruner1957.html.

[Buckland1991] Michael K. Buckland. 1991. *"Information as thing".* *Journal of the American Society for Information Science.* Vol. 42. no. 5. pp. 351-360.

http://onlinelibrary.wiley.com/doi/10.1002/%28SICI%291097-4571%28199106%2942:5%3C351::AID-ASI5%3E3.0.CO;2-3/abstract. doi: 10.1002/(SICI)1097-4571(199106)42:5<351::AID-ASI5>3.0.CO;2-3.

[Buckland1997] Michael K. Buckland. 1997. *"What Is a 'Document'?"*. *Journal of the American Society for Information Science*. Vol. 48. no. 9. pp. 804-809.

[Budanitsky2006] Alexander Budanitsky and Graeme Hirst. 2006. *"Evaluating WordNet-based Measures of Lexical Semantic Relatedness"*. *Computational Linguistics*. Vol. 32. no. 1. pp. 13-47.

[Buettcher2010] Stefan Buettcher, Charles L. A. Clarke, and Gordon V. Cormack. 2010. *Information Retrieval: Implementing and Evaluating Search Engines*. Cambridge, MA.

THE MIT PRESS.

[Buhrmester2007] Jason Buhrmester. 2007. *"NFL Films' Exhaustive Archive Is Rushing Into the Digital Age"*. *Wired*. September. http://www.wired.com/culture/lifestyle/magazine/15-10/ps_nfl.

[Bulmer1970] R. N. H. Bulmer. 1970. *"Which came first,* *the chicken or the egghead?".* *In Échanges et communications: mélanges offerts à Claude Lévi-Strauss à l'occasion de son 60ème anniversaire* . Jean Pouillon. pp. 1069-1091. Paris.

MOUTON & CO.

[Burke1993] Colin Burke. 1993. *Information and Secrecy: Vannevar Bush, Ultra, and the other Memex.* Metuchen, NJ.

SCARECROW PRESS.

[Burrell2015] Jenna Burrell. September 15, 2015. *"How the Machine 'Thinks:' Understanding Opacity in Machine Learning Algorithms".* http://ssrn.com/abstract=2660674 and http://dx.doi.org/10.2139/ssrn.2660674.

[Burrell2004] Jenna Burrell, Tim Brooke, and Richard Beckwith. 2004. *"Vineyard Computing: Sensor Networks in Agricultural Production". Pervasive Computing, IEEE.* Vol. 3. no. 1. pp. 38-45.

[Bush1945] Vannevar Bush. 1945. *"As We May Think". The Atlantic.* July.

[Buttcher2010] Stefan Büttcher, Charles Clarke, and Gordon V. Cormack.

2010. *Information retrieval: Implementing and evaluating search engines.*

THE MIT PRESS.

[Byrne2010] Gillian Byrne. 2010. *“The Strongest Link: Libraries and Linked Data”. D-Lib Magazine.* Vol. 16. no. 11/12. http://www.dlib.org/dlib/november10/byrne/11byrne.html.

C

[Cairo2012] Alberto Cairo. 2012. *The Functional Art: An introduction to information graphics and visualization.*

NEW RIDERS.

[Campbell2011] Joseph Keim Campbell, Michael O’Rourke, and Matthew H. Slater. 2011. *Carving Nature at Its Joints:*

Natural Kinds in Metaphysics and Science. Cambridge, MA.

A BRADFORD BOOK.

[Cano2005] Pedro Cano, Eloi Batlle, Emilia Gómez, Ro De C. T. Gomes, and Madeleine Bonnet. 2005. *"Audio Fingerprinting: Concepts And Applications". In Computational Intelligence for Modelling and Prediction.* Saman K. Halgamuge and Lipo Wang, (Eds.). New York.

SPRINGER.

[Carey1991] Susan Carey and Rochel Gelman. 1991. *The Epigenesis of Mind: Essays on Biology and Cognition.* New York.

PSYCHOLOGY PRESS.

http://www.amazon.co.uk/The-Epigenesis-Mind-Cognition-Symposia/dp/toc/0805804382.

[Carney2005] David Carney, David Fisher, Ed Morris, and Pat Place. 2005. *"Some Current Approaches to Interoperability". Carnegie Mellon Software Engineering Institute.*

[Carr2010] Patrick Carr. 2010. *"Forcing the moment to its crisis: Thoughts on pay-per-view and the perpetual* *access ideal”. Against the Grain*. Vol. 21. no. 6. pp. 16-18. http://hdl.handle.net/10342/2181.

[Carroll2010] Evan Carroll and John Romano. 2010. *Your Digital Afterlife: When Facebook, Flickr, and Twitter Are Your Estate, What’s Your Legacy?*. San Francisco, CA. NEW RIDERS.

[Casson2002] Lionel Casson. 2002. *Libraries in the Ancient World*. New Haven, CT. YALE UNIVERSITY PRESS.

[Cerf1969] Vint Cerf. 1969. *ASCII Format for Network Interchange*. http://tools.ietf.org/html/rfc20.

[Chaffin1984] Roger Chaffin and Douglas J. Herrmann. 1984. *“The similarity and diversity of semantic relations”. Memory & Cognition*. Vol. 12. no. 2. pp. 134-41.

[Chan2006] Lois Mai Chan and Marcia Lei Zeng. 2006. *“Metadata Interoperability and Standardization: A Study of*

Methodology Part I". D-Lib Magazine. Vol. 12. no. 2. http://www.dlib.org/dlib/june06/chan/06chan.html.

[Chandler1977] Alfred Dupont Chandler Jr. 1977. *The Visible Hand: The Managerial Revolution in American Business.* Cambridge, MA.

BELKNAP PRESS.

[Chandola2009] Varun Chandola, Arindam Banerjee, and Vipin Kumar. 2009. *"Anomaly detection: A survey.". ACM Computing Surveys (CSUR).* 41. no. 3. pp. 15.

[Chapman2009] Nigel Chapman and Jenny Chapman. 2009. *Digital Multimedia.* Hoboken, NJ.

WILEY.

[Chater2016] N Chater and G Loewenstein. 2016. *"The under-appreciated drive for sense-making.". Journal of Economic Behavior & Organization.*

[Chen2010] Donglin Chen, Xiaofei Li, Yueling Liang, and Jun Zhang. 2010. *"Research on the Theory of Customer-Oriented E-Catalog Ontology Automatic Construction".* *2010*

International Conference on E-Business and E-Government. pp. 2961-2964. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5590430. doi: 10.1109/ICEE.2010.748.

[Cherbakov2005] Luba Cherbakov, George Galambos, Ray Harishankar, Shankar Kalyana, and Guy Rackham. 2005. *"Impact of service orientation at the business level"*. *IBM Systems Journal*. Vol. 44. no. 4. pp. 653-668. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5386696. doi: 10.1147/sj.444.0653.

[Chi1981] M.T. Chi, P.J. Feltovich, and R. Glaser. 1981. *"Categorization and representation of physics problems by experts and novices"*. *Cognitive Science*. Vol. 5. 2. pp. 121-152.

[Cho2000] Junghoo Cho and Hector Garcia-Molina. 2000. *"The Evolution of the Web and Implications for an Incremental Crawler"*. *Proceedings of the 26th International Conference on Very Large Data Bases*. pp. 200-209.

[Chomsky1957] Noam Chomsky. 1957. *Syntactic Structures*. The Hague.

MOUTON & CO.

[Chomsky1965] Noam Chomsky. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA.

THE MIT PRESS.

http://www.worldcat.org/title/aspects-of-the-theory-of-syntax/oclc/309976.

[Christ1984] Karl Christ. 1984. *The handbook of medieval library history*. Metuchen, NJ.

SCARECROW PRESS.

[Christen2006] Peter Christen. 2006. *A Comparison of Personal Name Matching: Techniques and Practical Issues*. *The Australian National University*. (September). p. 14.

[Clark1868] Stephen Watkins Clark. 1868. *A Practical Grammar: In Which Words, Phrases, and Sentences Are Classified According to Their Offices and Their Various Relations to One Another*.

A.S. BARNES & CO.

http://archive.org/details/practicalgrammar00clar.

[Clark2010] Stephen Watkins Clark. 2010. *A Practical Grammar: In Which Words, Phrases, and Sentences Are Classified According to Their Offices and Their Various Relations to One Another.* Charleston, SC.

NABU PRESS. ORIGINALLY PUBLISHED IN 1847 BY A.S. BARNES & CO.

http://archive.org/details/practicalgrammar00clar.

[Coase1937] Ronald H. Coase. 1937. *"The Nature of the Firm".* *Economica, New Series.* Vol. 4. no. 16. pp. 386-405.

[Codd1970] E. F. Codd. 1970. *"A relational model of data for large shared data banks".* *Communications.* Vol. 13. no. 6. pp. 377-387. http://www.ncbi.nlm.nih.gov/pubmed/9617087.

[Conklin1987] Jeff Conklin. 1987. *"Hypertext : An Introduction and Survey".* *IEEE Computer.* Vol. 20. no. 9. pp. 17-41.

[Conklin1988] Jeff Conklin and Michael L. Begeman. 1988. *"glBIS :*

A Hypertext Tool for Exploratory Policy Discussion". *ACM Transactions on Information Systems.* Vol. 6. no. 4. pp. 303-331.

[Constantin1994] James A. Constantin and Robert F. Lusch. 1994. *Understanding Resource Management: How to deploy your people, products, and processes for maximum productivity.* Burr Ridge, IL.

IRWIN PROFESSIONAL.

[Cope2001] David Cope. 2001. *Virtual Music: Computer Synthesis of Musical Style.* Cambridge, MA.

THE MIT PRESS.

Ellingson, Ter. Transcription: Ethnomusicology. Edited by Helen Myers. Macmillan, 1992.
pp. 110-152.

[Cormen2009] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. 2009. *Introduction to Algorithms.* Cambridge, MA.

THE MIT PRESS.

[Cowan2004] John Cowan and Richard Tobin. 2004. *XML Information Set.* *Recommendation of the*

World Wide Web Consortium (W3C). http://www.w3.org/TR/xml-infoset/.

[Cox2007] Ingemar Cox, Matthew Miller, Jeffrey Bloom, Jessica Fridrich, and Ton Kalker. 2007. *Digital Watermarking and Steganography*. Burlington, MA.

MORGAN KAUFMANN.

[Cowen2015] Ron Cowen. 4 May 2015. *Ghostly Voices from Thomas Edison's Dolls can now be Heard.*

NEW YORK TIMES.

[Coyle2006] Karen Coyle. 2006. *"Identifiers: Unique, Persistent, Global"*. *The Journal of Academic Librarianship.* Vol. 34. no. 4. pp. 428-431.

[Coyle2010a] Karen Coyle. 2010. *"Library Data in a Modern Context"*. *Library Technology Reports.* Vol. 46. no. 1. pp. 5-13.

[Coyle2010b] Karen Coyle. 2010. *"RDA in RDF"*. *Library Technology Reports.* Vol. 46. no. 2. pp. 26-36.

[Crandall2006] R Crandall and C Pomerance. 2006. *"Prime numbers: a computational perspective"*. *Springer Science & Business Media.* 182.

[Crawford2012] Stephanie Crawford and Bernadette Johnson. 06 March 2012. *"How the Nest Learning Thermostat Works"*.

HOWSTUFFWORKS.COM.

[Croft2009] Bruce W. Croft, Donald Metzler, and Trevor Strohman. 2009. *Search Engines: Information Retrieval in Practice.* Boston, MA.

ADDISON WESLEY.

[Crow2010] David Crow. 2010. *Visible Signs: An Introduction to Semiotics in the Visual Arts.* London.

AVA PUBLISHING.

[Ctein2010] Ctein. 2010. *Digital Restoration from Start to Finish: How to Repair Old and Damaged Photographs.* Waltham, MA.

FOCAL PRESS.

[Curran2011] Brendan Curran. 2011. *"Champagne, Planes, and Veins". Unpublished term paper.*

UNIVERSITY OF CALIFORNIA BERKELEY.

[Cutter1876] Charles Cutter. 1876. *Rules for a printed* *dictionary catalogue. Issued as part 2 of Special report on public libraries, by the United States Education Bureau.* Washington, DC.

GOVERNMENT PRINTING OFFICE.

http://www.openlibrary.org/books/OL24156277M.

D

[Darnton2011] Robert Darnton. 2011. *"Google's Loss: The Public's Gain". The New York Review of Books.* April 28. http://www.nybooks.com/articles/archives/2011/apr/28/googles-loss-publics-gain.

[Das2002] Sajal K. Das, Diane J. Cook, Amiya Bhattacharya, Edwin O Heierman III, and Tze-Yun Lin. 2002. *"The role of prediction algorithms in the MavHome smart home architecture". IEEE Wireless Communications.* (December). pp. 77-84.

[Date2003] C.J. Date. 2003. *An Introduction to Database Systems.* Boston, MA.

ADDISON WESLEY.

[Datta2008] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z. Wang. 2008. *"Image retrieval: ideas, influences, and trends of the new age". ACM Computing Surveys.* Vol. 40. no. 2. doi: 10.1145/1348246.1348248.

[Deerwester1990] Scott Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman. 1990. *"Indexing by Latent Semantic Analysis". Journal of the American Society for Information Science.* Vol. 41. no. 6. pp. 391–407.

[deLeon2003] David de Léon. 2003. *"Actions, Artefacts, and Cognition: An Enthnography of Cooking". Lund University Cognitive Studies.* Vol. 104.

[Demartini2006] Gianluca Demartini and Stefano Mizzaro. 2006. *"A Classification of IR Effectiveness Metrics IR Metrics: A Survey and a Classification". In Advances in Information Retrieval: 28th European Conference on IR Research, ECIR 2006.* Mounia Lalmas, Andy MacFarlane, Stefan Rüger, Anastasios Tombros, Theodora Tsikrika, and Alexei Yavlinsky, (Eds.). 488-491. Berlin.

SPRINGER BERLIN HEIDELBERG.

[Denton2007] William Denton. 2007. *"FRBR and the History of Cataloging". In Understanding FRBR: What It Is and How It Will Affect Our Retrieval Tools.* Arlene G. Taylor, (Ed.). 35-57. Santa Barbara, CA.

LIBRARIES UNLIMITED.

[DeRose1989] Steven J. DeRose. 1989. *"Expanding the Notion of Links". Proceedings of the second annual ACM conference on Hypertext (HYPERTEXT '89).* (November). pp. 249-257.

[DeRose2010] Steven J. DeRose, Eve Maler, David Orchard, and Norman Walsh. 2010. *XML Linking Language (XLink). Recommendation of the World Wide Web Consortium (W3C).* http://www.w3.org/TR/xlink11/.

[Desoky2010] Ashraf Desoky. 2010. *MoSCoW Prioritisation.* http://certifications.groupsite.com/beta/discussion/topics/310632/messages.

[Deutscher2011] Guy Deutscher. 2011. *Through the Language Glass: Why the World*

Looks Different in Other Languages. London.

ARROW BOOKS.

[Dey2001] Anind K. Dey. 2001. *“Understanding and Using Context”.* *Personal and Ubiquitous Computing.* Vol. 5. no. 1. pp. 4-7.

[Diaz2005] Alejandro M. Diaz. 2005. *Through the Google Goggles: Sociopolitical Bias in Search Engine Design.* *Stanford University Program in Science, Technology and Society.*

[Ding2004] Li Ding, Tim Finin, Anupam Joshi, and Rong Pan. 2004. *“Swoogle: a search and metadata engine for the semantic web”.* *Proceedings of the 2004 Conference on Information and Knowledge Management.* pp. 652-659.

[Doctorow2001] Cory Doctorow. 2001. *Metacrap.* http://www.well.com/~doctorow/metacrap.htm.

[Domingos2015] Pedro Domingos. 2015. *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World.*

BASIC BOOKS.

[Donnellan1966] Keith S. Donnellan. 1966. *“Reference and Definite Descriptions”. The Philosophical Review.* Vol. 75. no. 3. pp. 281-304.

[Dorai2002] Chitra Dorai and Svetha Venkatesh. 2002. *“Bridging the Semantic Gap in Content Management Systems: Computational Media Aesthetics”. Media Computing.* pp. 1-9.

[Dougherty1985] Janet W. D. Dougherty and Charles M. Keller. 1985. *“Taskonomy: A practical approach to knowledge structures”. In Directions in cognitive anthropology.* Janet W. D. Dougherty. pp. 161-174. Urbana.

UNIVERSITY OF ILLINOIS PRESS.

[Drapeau2010] Mark Drapeau. 2010. *“The Three Phases of Government 2.0”. O’Reilly Radar.* May 10. http://radar.oreilly.com/2010/05/the-three-phases-of-government.html.

[Dumais2003] Susan Dumais. 2003. *“Data-driven approaches to information access”. Cognitive Science.* Vol. 27. no. 3. pp. 491-524. http://onlinelibrary.wiley.com/doi/10.1207/s15516709cog2703_7/abstract. doi: 10.1016/S0364-0213(03)00013-2.

[Durkheim1963] Emile Durkheim and Marcel Mauss. 1963. *Primitive classification*. Vol. 273.

UNIVERSITY OF CHICAGO PRESS.

[Durtschi2004] C. Durtschi, W. Hillison, and C. Pacini. 2004. *"The effective use of Benford's law to assist in detecting fraud in accounting data"*. *Journal of Forensic Accounting*. 5. 1. 17-34.

E

[Efron2011] Miles Efron. 2011. *"Information Search and Retrieval in Microblogs"*. *Journal of the American Society for Information Science and Technology*. Vol. 62. no. 6. pp. 996-1008. http://onlinelibrary.wiley.com/doi/10.1002/asi.21512/abstract. doi: 10.1002/asi.21512.

[Efthyvoulou2008] George Efthyvoulou. 2008. *"Alphabet Economics: The link between* *names and reputation". The Journal of Socio-Economics.* Vol. 37. no. 3. pp. 1266-1285. http://sheffield.academia.edu/GeorgiosEfthyvoulou/Papers/330894/Efthyvoulou_G._2008_._Alphabet_Economics_The_link_between_names_and_reputation._The_Journal_of_Socio-_Economics_37_3_1266-1285.

[Eliot1934] T.S. Eliot. 1934. *"Choruses from the rock.". The Rock. A Pageant Play.*

[Elliott2008] M Elliott, A Fremont, P Pantoja, and P Lurie. 2008. *"A New Method for Estimating Race/Ethnicity and Associated Disparities Where Administrative Records Lack Self-Reported Race/Ethnicity". Health Services Research.* 43. (5 Pt 1). pp 1722–1736. doi:10.1.

[Ellis2014] B Ellis. 2014. *Real-time analytics: Techniques to analyze and visualize streaming data.* Hoboken, NJ.

JOHN WILEY & SONS.

[Elman2009] Jeffrey L. Elman. 2009. *"On the meaning of words and dinosaur bones: Lexical* *knowledge without a lexicon”. Cognitive Science*. Vol. 33. no. 4. pp. 547-582. http://onlinelibrary.wiley.com/doi/10.1111/j.1551-6709.2009.01023.x/abstract. doi: 10.1111/j.1551-6709.2009.01023.x.

[Engelbart1963] Douglas Engelbart. 1963. *“A Conceptual Framework for the Augmentation of Man’s Intellect”. In Vistas in Information Handling*. Paul Howerton, (Ed.). 1-29. N.p.

SPARTAN BOOKS.

[Erl2005a] Thomas Erl. 2005. *Service-Oriented Architecture (SOA): Concepts, Technology, and Design*. Upper Saddle River, NJ.

PRENTICE HALL.

[Erl2005b] Thomas Erl. 2005. *Service-Oriented Architecture: A Field Guide to Integrating XML and Web Services*. Upper Saddle River, NJ.

PRENTICE HALL.

[Evans2000] G. Edward Evans and Margaret R. Zarnosky. 2000. *Developing Library and*

Information Center Collections. Santa Barbara, CA.

LIBRARIES UNLIMITED.

F

[Fabricant2002] Florence Fabricant. 2002. *"Chilean Sea Bass: More Than an Identity Problem"*. *The New York Times*. 2002-05-29.

[Falkenhainer1989] Brian Falkenhainer, Kenneth D. Forbus, and Dedre Gentner. 1989. *"The structure-mapping engine: Algorithm and examples."*. *Artificial intelligence*. 41. no. 1. pp. 1-63.

[Farish2002] J. Brian Farish. 2002. *What's in a Name?*. *Vertaasis*. http://www.vertaasis.com/articles/whats_in_a_name.htm.

[Feinberg2012] Melanie Feinberg. 2012. *"Synthetic Ethos: The Believability of Collections at* *the Intersection of Classification and Curation”*. *The Information Society*. Vol. 28. no. 5. pp. 329-339. http://www.tandfonline.com/doi/abs/10.1080/01972243.2012.708709. doi: 10.1080/01972243.2012.708709.

[Ferguson2002] K Ferguson. 2002. *Tycho and Kepler: the unlikely partnership that forever changed our understanding of the heavens.*. USA. BLOOMSBURY PUBLISHING.

[Fetterly2003] Dennis Fetterly, Mark Manasse, Mark Najork, and Janet Wiener. 2003. *“A Large-Scale Study of the Evolution of Web Pages”*. *Proceedings of the Twelfth International World Wide Web Conference*.

[Few2004] Stephen Few. 2004. *Show me the numbers: Designing tables and graphs to enlighten.* Vol. 1. no. 1. Oakland, CA. ANALYTICS PRESS.

[Few2012] Stephen Few. 2012. *Show Me the Numbers: Designing Tables and Graphs to Enlighten.* 2nd edition.

[Fidel2012] Raya Fidel. 2012. *Human Information Interaction: An Ecological Approach to Information Behavior.* Cambridge, MA.

THE MIT PRESS.

[Fillmore2000] Charles J. Fillmore and B. T. S. Atkins. 2000. *"Describing polysemy: The case of 'crawl'".* *In Polysemy: Theoretical and computational approaches*. Yael Ravin and Claudia Leacock, (Eds.). 91–110. Oxford.

OXFORD UNIVERSITY PRESS.

[Fishman2003] Charles Fishman. 2003. *"The Wal-Mart You Don't Know".* *Fast Company*. December 1. http://www.fastcompany.com/47593/wal-mart-you-dont-know.

[Fister2009] Barbara Fister. 2009. *"The Dewey Dilemma".* *Library Journal.*

[Flach2012] P. Flach. 2012. *Machine learning: the art and science of algorithms that make sense of data. .*

CAMBRIDGE UNIVERSITY PRESS.

[Florey2012] Kitty Burns Florey. 2012. *"A Picture of Language"*. *The New York Times*. 2012-03-26.

[Foer2011] Joshua Foer. 2011. *Moonwalking with Einstein: The art and science of remembering everything.*

PENGUIN.

[Freeman2005] Geoffrey T. Freeman, Scott Bennett, Sam Demas, Bernard Frischer, Christina A. Peterson, and Kathleen B. Oliver. 2005. *Library as Place: Rethinking Roles, Rethinking Space.* Washington, DC.

COUNCIL ON LIBRARY AND INFORMATION RESOURCES.

[Frege1892] Gottlob Frege. 1892. *"Uber Sinn und Bedeutung". In Zeitschrift fur Philosophie und philosophische Kritik 100.* pp. 25-50.

TRANSLATED AS "ON SENSE AND REFERENCE." IN TRANSLATIONS FROM THE PHILOSOPHICAL WRITINGS OF GOTTLOB FREGE, EDITED BY P.T. GEACH AND M. BLACK, 1952, PP. 56-78. OXFORD: BASIL BLACKWELL.

[Freitas2014] A.A. Freitas. 2014. *"Comprehensible classification models: a position paper.". ACM SIGKDD Explorations Newsletter, .* 15. no. 1. pp. 1-10.

[Fricke2009] Martin Fricke. 2009. *"The knowledge* *pyramid: a critique of the DIKW hierarchy".* *Journal of Information Science.* 35. no. 2. pp. 131-142.

[Friederici2009] Peter Friederici. March-April 2009. *Explaining Bird Flocks.*

AUDUBON.

https://www.audubon.org/magazine/march-april-2009/explaining-bird-flocks.

[Friedman1996] Batya Friedman and Helen Nissenbaum. 1996. *"Bias in Computer Systems".* *ACM Trans. Inf. Syst.*. 14. (3). p 330–47. . doi:10.1145/230538.230561.

[Fu2004] Xiang Fu, Tevfik Bultan, and Jianwen Su. 2004. *"Analysis of interacting BPEL web services".* *Proceedings of the 13th conference on World Wide Web—WWW '04.* doi: 10.1145/988672.988756.

[Furnas1987] G W Furnas, T K Landauer, L M Gomez, and S T Dumais. 1987. *"The Vocabulary Problem in Human-System Communication: an Analysis and a Solution".* *Communications of the ACM.* Vol. 30. no. 11. pp. 964-971.

[Furner2007] Jonathan Furner. 2007. *"User tagging of library resources: Toward a* *framework for system evaluation”. International Cataloguing and Bibliographic Control.* Vol. 37. pp. 47-51. http://archive.ifla.org/IV/ifla73/papers/157-Furner-en.pdf.

[Furner2008] Jonathan Furner. 2008. *“Interrogating ‘identity’: A philosophical approach to an enduring issue in knowledge organization”. Knowledge Organization.* Vol. 36. pp. 3-16.

G

[Garfield2000] Eugene Garfield, Helen Barsky Atkins, and Blaise Cronin. 2000. *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield.* Medford, NJ.

INFORMATION TODAY INC.

[Gaukroger2001] Stephen Gaukroger. 2001. *Francis Bacon* *and the Transformation of Early-Modern Philosophy*. Cambridge, UK. CAMBRIDGE UNIVERSITY PRESS.

[Geller1999] Jacklyn Geller. 1999. *“The Contemporary Wedding Invitation: A Social Document in Crisis”*. *Salmagundi*. Vol. 121/122. pp. 175-187.

[Geller2012] Tom Geller. 2012. *“Talking to machines”*. *Communications of the ACM*. 55. no. 4. pp 14-16.

[Gentner1983] Dedre Gentner. 1983. *“Structure-mapping: A theoretical framework for analogy”*. *Cognitive Science*. Vol. 7. no. 2. pp. 155-170. http://onlinelibrary.wiley.com/doi/10.1207/s15516709cog0702_3/abstract. doi: 10.1016/S0364-0213(83)80009-3.

[Gentner1997] D. Gentner, S. Brem, R. W. Ferguson, A.B. Markman, B.B. Levidow, P. Wolff, and K. D Forbus, . 1997. *“Analogical reasoning and conceptual change: A case study of Johannes Kepler”*. *The journal of the learning sciences*. 6. no. 1. pp. 3-40.

[Gershenfeld2004] Neil Gershenfeld, Raffi Krikorian, and Danny Cohen. 2004. *“The Internet of Things”*. *Scientific American*. September 27.

http://www.scientificamerican.com/article.cfm?id=the-internet-of-things.

[Getty2006] . 2006. *CDWA Lite: Specification for an XML Schema for Contributing Records via the OAI Harvesting Protocol.* http://getty.edu/research/publications/electronic_publications/cdwa/cdwalite.pdf.

[Gibson1977] James J. Gibson. 1977. *“The Theory of Affordances”.* *In Perceiving, Acting, and Knowing: Toward an Ecological Psychology.* Robert Shaw and John Bransford, (Eds.). Mahwah, NJ.

LAWRENCE ERLBAUM ASSOCIATES.

[Gill2008] Tony Gill, Anne J. Gilliland, Maureen Whalen, and Mary S. Woodly. 2008. *Introduction to Metadata.* Los Angeles, CA.

GETTY PUBLICATIONS.

http://www.getty.edu/research/publications/electronic_publications/intrometadata/index.html.

[Gillies2000] James Gillies and Robert Cailliau. 2000. *How the Web Was Born: The story of the World Wide Web.* New York.

OXFORD UNIVERSITY PRESS, USA.

[Gilliland-Swetland2000] AJ Gilliland-Swetland. 2000. *“Enduring paradigm, new opportunities: the value of the archival perspective in the digital environment”. Communications.* February.

[Giridharadas2014] Anand Giridharadas. 6 February 2014. *“Seeing the Show With Their Ears”.* NEW YORK TIMES.

[Gjerdingen2007] Robert O. Gjerdingen. 2007. *Music in the Galant Style.* OXFORD UNIVERSITY PRESS.

[Gladwell1996] Malcolm Gladwell. 1996. *“The Science of Shopping”. The New Yorker.* November 4.

[Glushko1988] Robert J. Glushko, Mark D. Weaver, Thomas A. Coonan, and Janet E. Lincoln. 1988. *“Hypertext engineering: practical methods for creating a compact disk encyclopedia”. Proceedings of the ACM Conference on Document Processing Systems—DOCPROCS ’88.* pp. 11-19.

[Glushko2005] Robert J. Glushko and Tim McGrath. 2005. *Document Engineering: Analyzing* *and Designing Documents for Business Informatics and Web Services*. Cambridge, MA.

THE MIT PRESS.

http://mitpress.mit.edu/catalog/item/default.asp?ttype=2&tid=10476.

[Glushko2008] Robert J. Glushko, Paul P. Maglio, Teenie Matlock, and Lawrence W. Barsalou. 2008. "Categorization in the wild". *Trends in cognitive sciences*. Vol. 12. no. 4. pp. 129-35. http://www.ncbi.nlm.nih.gov/pubmed/18343710. doi: 10.1016/j.tics.2008.01.007.

[Glushko2013] Robert J. Glushko and Karen J. Nomorosa. 2013. "Substituting Information for Interaction: A Framework for Personalization in Service Encounters and Service Systems". *Journal of Service Research*. Vol. 16. no. 1. pp. 21-38. doi: 10.1177/1094670512463967.

[Glushko2013b] Robert J. Glushko. 2013. *The Discipline of Organizing*.

THE MIT PRESS.

[Glushko2015] Robert J. Glushko. 2015. *Collaborative Authoring, Evolution, and Personalization for a "Transdisciplinary" Textbook. OpenSym'15*. http://dx.doi.org/10.1145/2789853.2789867.

[Godby2008] Carol Jean Godby, Devon Smith, and Eric Childress. 2008. *"Toward element-level interoperability in bibliographic metadata". Code{4}lib*. no. 2.

[Goldberg2008] Kevin Howard Goldberg. 2008. *XML: Visual QuickStart Guide*. San Francisco, CA.

PEACHPIT PRESS.

[Goldstone2009] R.L. Goldstone and T. M. Gureckis. 2009. *"Collective behavior.". Topics in cognitive science*. 1. 3. p 412-438.

[Golder2006] Scott A. Golder and Bernardo A. Huberman. 2006. *"Usage patterns of collaborative tagging systems.". Journal of Information Science*. 32. 2. 198-208.

[Goldstone1994] R L Goldstone. 1994. *"The role of similarity in categorization:*

providing a groundwork". *Cognition*. Vol. 52. no. 2. pp. 125-57. http://www.ncbi.nlm.nih.gov/pubmed/7924201.

[Goodwin2012] Richard Goodwin, Sweefen Goh, Pietro Mazzolen, Vibha Sinha, Debdoot Mukherjee, and Senthil Mani. 2012. *"Effective Content Reuse for Business Consulting Practices"*. *2012 Annual SRII Global Conference*. pp. 682-690. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6311054. doi: 10.1109/SRII.2012.82.

[Gorman2004] Michael Gorman. 2004. *The Concise AACR2*. Chicago.

AMERICAN LIBRARY ASSOCIATION EDITIONS.

[Grainger1908] Percy Grainger. 1908. *Collecting with the Phonograph*. *Journal of the Folk Song Society*. 12. pp 147-242.

[Grappone2011] Jennifer Grappone and Gravida Couzin. 2011. *Search Engine Optimization (SEO): An Hour a Day*. Hoboken, NJ.

SYBEX.

[Gravois2010] John Gravois. July/August 2010. *"The Agnostic Cartographer"*. *Washington Monthly*.

http://www.washingtonmonthly.com/features/2010/1007.gravois.html.

[Grean2005] Michael Grean and Michael J. Shaw. 2005. *"Supply-Chain Integration through Information Sharing: Channel Partnership between Wal-Mart and Procter & Gamble". Center for IT and e-Business Management, University of Illinois at Urbana-Champaign.*

[Grimmelmann2009] James Grimmelmann. 2009. *"How to Fix the Google Book Search Settlement". Journal of Internet Law.* April. http://works.bepress.com/james_grimmelmann/23/.

[Gross1990] Derek Gross and Katherine J. Miller. 1990. *"Adjectives in WordNet". International Journal of Lexicography.* Vol. 3. no. 4. pp. 265-277. doi: 10.1093/ijl/3.4.265.

[Gruber1993] Thomas R. Gruber. 1993. *"A Translation Approach to Portable Ontology Specifications". Knowledge Acquisition.* Vol. 5. no. 2. pp. 199-220.

[Grudin1994] Jonathan Grudin. 1994. *"Groupware and social dynamics: eight challenges* *for developers"*. *Communications of the ACM*. Vol. 37. no. 1. pp. 92-105.

[Guarino1998] Nicola Guarino. 1998. *"Formal Ontology and Information Systems"*. *Formal ontology in information systems: proceedings of FOIS '98*. (June). pp. 3-15.

[Guenther2009] Rebecca Guenther and Robert Wolfe. 2009. *"Integrating Metadata Standards to Support Long-Term Preservation of Digital Assets: Developing Best Practices for Expressing Preservation Metadata in a Container Format"*. *iPRES 2009: the Sixth International Conference on the Preservation of Digital Objects, California Digital Library, UC Office of the President*.

[Guijarro2007] Luis Guijarro. 2007. *"Interoperability frameworks and enterprise architectures in e-government initiatives in Europe and the United States"*. *Government Information Quarterly*. Vol. 24. no. 1. pp. 89-101.

http://linkinghub.elsevier.com/retrieve/pii/S0740624X06000864. doi: 10.1016/j.giq.2006.05.003.

H

[Halasz1994] Frank Halasz and Mayer Schwartz. 1994. *"The Dexter Hypertext Reference Model"*. *Communications of the ACM*. Vol. 37. no. 2. pp. 30-39.

[Hall2010] Jeffrey A. Hall, Namkee Park, Hayeon Song, and Michael J. Cody. 2010. *"Strategic misrepresentation in online dating: The effects of gender, self-monitoring, and personality traits"*. *Journal of Social and Personal Relationships*. 27. no. 1. pp 117-135.

[Halvey2007] Martin Halvey and Mark T. Keane. 2007. *"An Assessment of Tag Presentation Techniques"*. *Proceedings of the 16th* *international conference on World Wide Web (WWW '07)*. pp. 1313-1314.

[Halvorson2012] Kristina Halvorson and Melissa Rach. 2012. *Content strategy for the web.* NEW RIDERS.

[Hamilton2012] Kate Hamilton and Lauren Wood. 2012. *"Schematron in the Context of the Clinical Document Architecture (CDA)". Proceedings of Balisage: The Markup Conference 2012*. Vol. 8. doi: 10.4242/BalisageVol8.Wood01.

[Hammond2004] Tony Hammond, Timo Hannay, Ben Lund, and Joanna Scott. 2004. *"Social Bookmarking Tools (I): A General Review". D-Lib Magazine*. Vol. 11. no. 4. http://www.dlib.org/dlib/april05/hammond/04hammond.html.

[Hanington2012] B. Hanington and B. Martin. 2012. *Universal Methods of Design: 100 Ways to Research Complex Problems, Develop Innovative Ideas, and Design Effective Solutions.* Beverly, MA. ROCKPORT PUBLISHERS.

[Hansen2009] Morten Hansen. *Collaboration: How* *leaders avoid the traps, build common ground, and reap big results.* HARVARD BUSINESS PRESS. 2009.

[Harpring2009] Patricia Harpring. 2009. *Categories for the Description of Works of Art. Getty Research Institute.* http://getty.edu/research/publications/electronic_publications/cdwa/.

[Harris1996] Roy Harris. 1996. *Signs of Writing.* London. ROUTLEDGE.

[Haspelmath2010] Martin Haspelmath and Andrea Sims. 2010. *Understanding Morphology.* London. ROUTLEDGE.

[Haviland1998] John B. Haviland. 1998. *"Guugu Yimithirr Cardinal Directions". Ethos.* Vol. 26. no. 1. pp. 25-47. doi: 10.1525/eth.1998.26.1.25.

[He2007] Bin He, Mitesh Patel, Zhen Zhang, and Kevin Chen-chuan Chang. 2007. *"Accessing the Deep Web: A Survey". Communications of the ACM.* Vol. 50. no. 5. pp. 94-101.

[Hearst2009] Marti A. Hearst. *Search User Interfaces.*

CAMBRIDGE UNIVERSITY PRESS.

[Heath2011] Tom Heath and Christian Bizer. 2011. *Linked Data: Evolving the Web into a Global Data Space.* San Rafael, CA.

MORGAN & CLAYPOOL PUBLISHERS.

[Heidorn2008] P. Bryan Heidorn. 2008. *Shedding Light on the Dark Data in the Long Tail of Science. Library Trends, Volume 57, Number 2, Fall 2008, pp. 280-299 (Article).*

THE JOHNS HOPKINS UNIVERSITY PRESS.

doi: 10.1353/lib.0.0036.

[Heller2012] Daphna Heller, Kristen S. Gorman, and Michael K. Tanenhaus. 2012. *"To name or to describe: shared knowledge affects referential form". Topics in cognitive science.* Vol. 4. no. 2. pp. 290-305. http://www.ncbi.nlm.nih.gov/pubmed/22389094.

[Helper2003] Susan Helper and John Paul MacDuffie. 2003. *B2B and modes of exchange:*

evolutionary and transformative effects. *The Global Internet Economy.* . 331-380.

[Hemerly2011] Jess Hemerly. 2011. *"Making Metadata: The Case of MusicBrainz"*. *SSRN Electronic Journal.* http://papers.ssrn.com/abstract=1982823.

[Herbst2009] Charles Herbst. 2009. *"Death in cyberspace"*. *Res Gestae.* Vol. 53. no. 10. pp. 16-25.

[Hillmann2005] Diane Hillmann. 2005. *Using Dublin Core—The Elements*. *Dublin Core Metadata Initiative.* http://dublincore.org/documents/usageguide/elements.shtml.

[Hjorland1992] Birger Hjørland. 1992. *"The Concept of 'Subject' in Information Science"*. *Journal of Documentation.* Vol. 48. no. 2. pp. 172-200. http://www.emeraldinsight.com/journals.htm?issn=0022-0418&volume=48&issue=2&articleid=1650137&show=html.

[Hjorland2001] Birger Hjørland. 2001. *"Why is meta analysis neglected by information scientists?"*. *Journal of the American Society for*

Information Science and Technology. Vol. 52. no. 13. pp. 1193-1194. doi: 10.1002/asi.1179.

[Hoffman2008] Michael Hoffman. 2008. *"Details emerging on how fuses got to Taiwan"*. *Air Force Times*. 2008-03-26. `http://www.airforcetimes.com/news/2008/03/airforce_loose_fuses3_032708w/`.

[Hofweber2009] Thomas Hofweber. 2009. *"Logic and Ontology"*. *In The Stanford Encyclopedia of Philosophy*. Edward N. Zalta, (Ed.). `http://plato.stanford.edu/archives/spr2009/entries/logic-ontology/`.

[Holman2001] G. Ken Holman. 2001. *Definitive XSLT and XPath*. Upper Saddle River, NJ.

PRENTICE HALL.

[Hori2010] Mitsuyoshi Hori, Eiji Kawashima, and Tomihoro Yamazaki. 2010. *"Application of Cloud Computing to Agriculture and Prospects in Other Fields"*. *Fujitsu Scientific and Technical Journal*. Vol. 46. no. 4. pp. 446-454.

[Howard2011] Jennifer Howard. 2011. *"Librarians Puzzle Over E-Books They May Buy but Not Truly Own"*. *The Chronicle of Higher*

Education. May 15. http://chronicle.com/article/Hot-Type-Librarians-Puzzle/127538/.

[Howe2006] Jeff Howe. 2006. *"The Rise of Crowdsourcing"*. *Wired*. June. http://www.wired.com/wired/archive/14.06/crowds.html.

[Howe2008] Jeff Howe. 2008. *Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business*. New York.

RANDOM HOUSE.

[Hu2004] Minqing Hu and Bing Liu. 2004. *"Mining and summarizing customer reviews"*. *Proceedings of the 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. pp. 168.

[Huang2014] Xuedong Huang, James Baker, and Raj Reddy. 2014. *"A historical perspective of speech recognition"*. *Communications of the ACM*. 57. no. 1. pp 94-103.

[Hutchins1977] W. John Hutchins. 1977. *"On the problem of 'aboutness' in document analysis."*. *Journal of Informatics*. 1. no. 1. 17-35.

[Hutchins2010] Edwin Hutchins. 2010. *"Cognitive Ecology".* *Topics in cognitive science.* Vol. 2. no. 4. pp. 705-715.
http://philpapers.org/rec/EDWCE.

[Hyvonen2004] Eero Hyvönen, Miikka Junnila, Suvi Kettula, Eetu Mäkelä, Samppa Saarela, Mirva Salminen, Ahti Syreeni, Arttu Valo, and Kim Viljanen. 2004. *"Finnish Museums on the Semantic Web: The User's Perspective on MuseumFinland".* *Museums and the Web 2004: International Conference for Culture and Heritage On-line.*
http://www.museumsandtheweb.com/mw2004/papers/hyvonen/hyvonen.html.

I

[Iyengar2000] Sheena S. Iyengar and Mark R. Lepper. 2000. *"When choice is demotivating: Can one desire too much of a good thing?".*

Journal of personality and social psychology. 79. . no. 6. . 995.

J

[Jackendoff1996] Ray S. Jackendoff. 1996. *The Architecture of the Language Faculty.* Cambridge, MA.

THE MIT PRESS.

[Jackman2015] T. Jackman. 10 January 2015. *As Fairfax library collection shrinks, citizen groups demand audits, more funding.*

WASHINGTON POST.

http://wpo.st/N2-N0.

[Jacobs2004] Ian Jacobs and Norman Walsh. 2004. *Architecture of the World Wide Web.* Vol. 1.

W3C.

Retrieved from http://www.w3.org/TR/webarch/.

[Jennex2009] Murray E. Jennex. 2009. *"Re-visiting the knowledge pyramid". HICSS'09. IEEE.* pp. 1-7.

[John2006] Ajita John and Dorée Seligmann. 2006. *"Collaborative Tagging and Expertise in the Enterprise". Proceedings of the Fifteenth International World Wide Web Conference.*

[John1999] Nathan John. 1999. *Sony: The private life.* HOUGHTON MIFFLIN.

[Johnson2013] Jeff Johnson. 2013. *Designing with the Mind in Mind: Simple Guide to Understanding User Interface Design Guidelines.* ELSEVIER.

[Johnson2010] Marilyn Johnson. 2010. *"U.S public libraries: We lose them at our peril". The Los Angeles Times.* 2010-07-06. http://articles.latimes.com/2010/jul/06/opinion/la-oe-johnson-libraries-20100706.

[Johnson1998] Spencer Johnson. 2010. *Who Moved My*

Cheese?. *An Amazing Way to Deal with Change in Your Work and in Your Life*. 1998. New York.

PUTNAM ADULT.

[Jones2007] William Jones. 2007. *Keeping Things Found: The Study and Practice of Personal Information Management*. Burlington, MA.

MORGAN KAUFMANN.

[Juola2014] Patrick Juola. 2014-08-10. *“Rowling and ‘Galbraith’: an authorial analysis”*. *Language Log*. http://languagelog.ldc.upenn.edu/nll/?p=5315.

[Juran1951] , Juran. 1951. *Quality Control Handbook*. New York.

MCGRAW-HILL.

K

[Kahn1995] R. Kahn and R. Wilensky. 1995. *A Framework for Distributed Digital Object Services*. Reston.

CORPORATION FOR NATIONAL RESEARCH INITIATIVES.

Retrieved from http://www.cnri.reston.va.us/k-w.html.

[Kahneman2003] Daniel Kahneman. 2003. *"Maps of bounded rationality: Psychology for behavioral economics"*. *American economic review*. 1449-1475.

[Kahneman1979] Daniel Kahneman and Amos Tversky. 1979. *"Prospect theory: An analysis of decision under risk"*. *Econometrica: Journal of the Econometric Society*. 263-291.

[Kalbach2007] James Kalbach. 2007. *Designing Web Navigation*.

O'REILLY MEDIA, INC.

[Kaser2007] Owen Kaser and Daniel Lemire. 2007. *"Tag-Cloud Drawing : Algorithms for Cloud*

Visualization”. WWW’07 Workshop on Taggings and Metadata for Social Information Organization.

[Kaynak1997] Erdener Kaynak and Paul Herbig. 1997. *Handbook of Cross-Cultural Marketing.* London.

ROUTLEDGE.

[Kent2012] William Kent and Steve Hoberman. 2012. *Data and Reality: A Timeless Perspective on Perceiving and Managing Information in Our Imprecise World.* Westfield, NJ.

TECHNICS PUBLICATIONS, LLC.

[Kesan2006] Jay P. Kesan and Rajiv C. Shah. 2006. *“Setting software defaults: Perspectives from law, computer science and behavioral economics”. Notre Dame L. Rev.*. 82. p. 583.

[Kilgour1998] Frederick G. Kilgour. 1998. *The Evolution of the Book.* Oxford.

OXFORD UNIVERSITY PRESS.

[Kim2009] Kenneth Kim, John R. Nofsinger, and Derek J. Mohr. 2009. *Corporate Governance*. Upper Saddle River, NJ.

PRENTICE HALL.

[Kim2003] W. Kim, B. J. Choi, E. K. Hong, S. K. Kim, and D. Lee. 2003. *A taxonomy of dirty data.*. *Data mining and knowledge discovery*. 7. 1. 81-99.

[Kimber2012] Eliot Kimber. 2012. *DITA for Practitioners Volume 1: Architecture and Technology*. Laguna Hills, CA.

XML PRESS.

[Kirsh1995] David Kirsh. 1995. *“The Intelligent Use of Space”*. *Artificial Intelligence*. Vol. 73. (1-2). pp. 31-36. doi: 10.1016/0004-3702(94)00017-U.

[Kirsh2000] David Kirsh. 2000. *“A Few Thoughts on Cognitive Overload”*. *Intellectica*. Vol. 30. no. 1. pp. 19-51.

[Kirschenbaum2008] Matthew G. Kirschenbaum. 2008. *Mechanisms: New Media and the Forensic Imagination*. Cambridge, MA.

THE MIT PRESS.

[Kirschenbaum2009] Matthew G. Kirschenbaum, Erika L. Farr, Kari M. Kraus, Naomi Nelson, Catherine Stollar Peters, and Gabriela Redwine.

2009. *“Digital Materiality: Preserving Access to Computers as Complete Environments”.* *iPRES 2009: The Sixth International Conference on the Preservation of Digital Objects, California Digital Library, UC Office of the President.* http://escholarship.org/uc/item/7d3465vg.

[Kleppe2003] Anneke Kleppe, Jos Warmer, and Wim Bast. 2003. *MDA Explained: The Model Driven Architecture: Practice and Promise.* Boston, MA.

ADDISON WESLEY.

[Koffka1935] Kurt Koffka. 1935. *Principles of Gestalt psychology.* 44.

ROUTLEDGE.

[Kondo2014] Marie Kondo. 2014. *The Life-changing Magic of Tidying Up.* *The Japanese Art of Decluttering and Organizing.*

TEN SPEED PRESS.

[Kulish2009] Nicholas Kulish. 2009. *“High Court in Germany Pops Names That Balloon”.* *The New York Times.* 2009-05-06.

http://www.nytimes.com/2009/05/06/world/europe/06germany.html.

[Kuniavsky2010] Mike Kuniavsky. 2010. *Smart Things: Ubiquitous Computing User Experience Design*. Burlington, MA.

MORGAN KAUFMANN.

[Kurosawa1957] Akira Kurosawa. 1957. *Kumonosu-jo (蜘蛛巣城). (Throne of Blood)*. Tokyo.

TOHO STUDIOS.

L

[Lackey1999] Douglas P. Lackey. 1999. *“What are the modern classics? The Baruch poll of great philosophy in the twentieth century”. Philosophical Forum.* Vol. 30. no. 4. pp. 329-346. http://philpapers.org/rec/LACWAT.

[Lagoze1996] Carl Lagoze. 1996. *A Container Architecture for Diverse Sets of Metadata. D-Lib Magazine.* 2(7/8). Retrieved from http://www.dlib.org/dlib/july96/lagoze/07lagoze.html.

[Lagoze2005] Carl Lagoze, S. Payette, E. Shin, and C. Wilper. 2005. *Fedora: An architecture for complex objects and their relationships. International Journal of Digital Libraries.* 6(2). 124–138.

[Lagoze2008] Carl Lagoze, H. Van De Sompel, M.L. Nelson, S. Warner, R. Sanderson, and P. Johnston. 2008. *Object Re-Use & Exchange: A Resource-Centric Approach (arXiv).* Retrieved from `http://arxiv.org/abs/0804.2273`.

[Lakoff1990] George Lakoff. 1990. *Women, Fire, and Dangerous Things.* Chicago.

UNIVERSITY OF CHICAGO PRESS.

[Lancaster1968] Frederick Wilfrid Lancaster. 1968. *Information Retrieval Systems: Characteristics, Testing, and Evaluation.* Hoboken, NJ.

JOHN WILEY & SONS.

[Langevoort2006] Donald Langevoort. 2006. *“Internal Controls After Sarbanes-Oxley: Revisiting Corporate Law’s ‘Duty of Care as Responsibility for Systems’”. Georgetown Law Faculty Publications and Other Works.* http://scholarship.law.georgetown.edu/facpub/144.

[Langville2012] Amy Langville and Carl Meyer. 2012. *Google’s Page Rank and Beyond: The Science of Search Engine Rankings.* Princeton, NJ.

PRINCETON UNIVERSITY PRESS.

[Larose2014] Daniel T. Larose. 2014. *Discovering knowledge in data: an introduction to data mining..*

JOHN WILEY & SONS.

[Laskey2005] Kenneth J. Laskey. 2005. *“Metadata Concepts to Support a Net-Centric Data Environment”. In Net-Centric Approaches to Intelligence and National Security.* Roy Ladner and Frederick E. Petry, (Eds.). 29-54. New York.

SPRINGER.

[LeCun2015] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. *“Deep learning.”*. *Nature*. 521. no. 7553. pp. 436-444.

[Lee1993] Sharon M. Lee. 1993. *“Racial classifications in the US census: 1890-1990”*. *Ethnic and Racial Studies*. Vol. 16. no. 1. pp. 75-94.

[Lee2007] John A. Lee and Michel Verleysen. 2007. *Nonlinear Dimensionality Reduction*. New York.

SPRINGER.

[Leonardi2010] Paul M. Leonardi. 2010. *“Digital materiality? How artifacts without matter, matter”*. *First Monday*. Vol. 15. (6-7). http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/3036/2567.

[Levitin2002] Daniel J. Levitin. 2002. *Foundations of Cognitive Psychology: Core Readings*. Cambridge, MA.

A BRADFORD BOOK.

[Levitin2014] Daniel J. Levitin. 2014. *The organized mind: Thinking straight in the age of information overload.*.

PENGUIN.

[Levitt2005] Steven D. Levitt and Stephen J. Lubner. 2005. *Freakonomics.* Vol. 61.

WILLIAM MORROW.

[Levy2006] Robert L. Levy and Patricia L. Casey. 2006. *Electronic Evidence and the Large Document Case: Common Evidence Problems. Haynes and Boone, LLP.* July.

[Levy2010] Steven Levy. 2010. *"How Google's Algorithm Rules the Web". Wired.* February 22. http://www.wired.com/magazine/2010/02/ff_google_algorithm/.

[Lewis2003] Michael Lewis. 2003. *Moneyball: The art of winning an unfair game.*

WW NORTON & COMPANY.

[Lewis2010] Michael Lewis. 2010. *The Big Short: Inside the Doomsday Machine..*

WW NORTON & COMPANY.

[Lidz2003] Franz Lidz. 2003. *Ghosty Men: The Strange But True Story of the Collyer Brothers, New York's*

Greatest Hoarders: an Urban Historical.

BLOOMSBURY PUBLISHING.

[Linhares2007] M.T. Linhares and P. Brum. 2007. *Understanding our understanding of strategic scenarios: What role do chunks play?.* *Cognitive Science.* Vol. 31. 6. pp. 989-1007.

[Linnaeus1735] C. von Linnaeus. 1735. *Systema Naturae 1.* Editio Decima, Reformata.

(HOLMIAE, FASC REPRINT 1939) (1758).

Bri Mus Nat Hist, London.

[Linthicum1999] David S. Linthicum. 1999. *Enterprise Application Integration.* Boston, MA.

ADDISON WESLEY.

[Liptak2014] Adam Liptak. *Justices to weigh finance law as it was applied to little fish.* April 28 2014.

NY TIMES.

http://www.nytimes.com/2014/04/29/us/politics/justices-to-weigh-fishermans-conviction-under-federal-finance-law.html.

[Lockyer1893] Norman Lockyer. 1893. *The Dawn of astronomy: a study of the temple-worship and mythology of the ancient Egyptians.*

MACMILLAN AND COMPANY.

[Lohr2014] Steve Lohr. 2014. *“For big-data scientists, janitor work is key hurdle to insights.”.* 17.

THE NEW YORK TIMES.

[Lorch1989] Robert F. Lorch. 1989. *“Text-signaling devices and their effects on reading and memory processes”.* *Educational Psychology Review.* Vol. 1. no. 3. pp. 209-234. doi: 10.1007/BF01320135.

[Loshin2008] David Loshin. 2008. *Master Data Management.* Burlington, MA.

MORGAN KAUFMANN.

[Lovink2011] Geert Lovink and Nathaniel Tkacz. 2011. *Critical Point of View: A Wikipedia Reader.* Amsterdam.

INSTITUTE OF NETWORK CULTURES.

http://www.amazon.co.uk/CRITICAL-POINT-OF-VIEW-WIKIPEDIA/dp/9078146133.

[Lubetzky2001] Seymour Lubetzky. 1953. *Seymour Lubetzky: Writings on the Classical Art of Cataloging.* Santa Barbara, CA.

LIBRARIES UNLIMITED.

M

[Madrigal2009] Alexis Madrigal. 2009. *"Autonomous Robots Invade Retail Warehouses". Wired.* January 27. http://www.wired.com/wiredscience/2009/01/retailrobots/.

[Madrigal2014] Alexis Madrigal. 2014. *"How Netflix Reverse Engineered Hollywood".*

THE ATLANTIC.

http://www.theatlantic.com/technology/archive/2014/01/how-netflix-reverse-engineered-hollywood/282679/.

[Maglio2009] Paul P. Maglio, Stephen L. Vargo, Nathan Caswell, and Jim Spohrer. 2009. *"The service system is the basic abstraction of service science."*. *Information Systems and e-business Management*. 7. no. 4. 395-406.

[Malaga2008] Ross A. Malaga. 2008. *"Worst Practices in Search Engine Optimization"*. *Communications of the ACM*. Vol. 51. no. 12. pp. 147-150. doi: 10.1145/1409360.1409388.

[Malone1983] Thomas W. Malone. 1983. *"How do people organize their desks?: Implications for the design of office information systems"*. *ACM Transactions on Information Systems*. Vol. 1. no. 1. pp. 99-112.

[Malt1995] Barbara C. Malt. 1995. *"Category Coherence in Cross-Cultural Perspective"*. *Cognitive Psychology*. Vol. 29. no. 2. pp. 85-148. http://www.eric.ed.gov/ERICWebPortal/detail?accno=EJ514283.

[Maness2006] Jack M. Maness. 2006. *"Library 2.0 Theory: Web 2.0 and its Implications for Libraries"*. *Webology*.

Vol. 3. no. 2. http://www.webology.org/2006/v3n2/a25.html.

[Mann1988] William C. Mann and Sandra A. Thompson. 1988. *“Rhetorical Structure Theory: A Theory of Text Organization”*. *Text*. Vol. 8. no. 3. pp. 243-281.

[Manning2008] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. 2008. *Introduction to Information Retrieval*. Cambridge, UK. CAMBRIDGE UNIVERSITY PRESS.

[Manyika2011] James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers. 2011. *“Big data: The next frontier for innovation, competition, and productivity”*. *McKinsey Global Institute*. http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation.

[Marchioni2012] G. Marchioni, C. Lee, H. Bowden, and M. Lesk. 2012. *“Curating for quality: Ensuring data quality to enable new science”*. *Final Report of NSF Workshop*. http://datacuration.web.unc.edu/files/2012/10/NSF_Data_Curation_Workshop_Report.pdf
.

[Marcotte2011] Ethan Marcotte. 2011. *Responsive web design.*

EDITIONS EYROLLES.

[Mardia1980] Kanti V. Mardia, J. T. Kent, and J. M. Bibby. 1980. *Multivariate Analysis.* Waltham, MA.

ACADEMIC PRESS.

[Margolis1999] Eric Margolis and Stephen Laurence. 1999. *Concepts: Core Readings.* Cambridge, MA.

A BRADFORD BOOK.

[Markoff2011] John Markoff and Somini Sengupta. 2011. *“Separating you and me? 4.74 degrees”.* *The New York Times*. 2011.

[Marlow2006] Cameron Marlow, Mor Naaman, Danah Boyd, and Marc Davis. 2006. *“HT06, tagging paper, taxonomy, Flickr, academic article, to read.”.* *In Proceedings of the Seventeenth Conference on Hypertext and Hypermedia*. 31-40.

ACM.

[Maron1977] M. E. Maron. 1977. *“On indexing, retrieval and the meaning of about”.* *Journal of the American Society for*

Information Science. Vol. 28. no. 1. pp. 38-43. doi: 10.1002/asi.4630280107.

[Marsh2006] Jonathan Marsh, David Orchard, and Daniel Veillard. 2006. *XML Inclusions (XInclude). Recommendation of the World Wide Web Consortium (W3C)*. http://www.w3.org/TR/xinclude/.

[Marshall2007] Catherine C. Marshall, Frank McCown, and ML Nelson. 2007. *"Evaluating personal archiving strategies for Internet-based information". Proceedings of IS&T Archiving 2007*. http://arxiv.org/abs/0704.3647.

[Marshall2008] Catherine C. Marshall. 2008. *"Rethinking Personal Digital Archiving Part 1: Four Challenges from the Field". D-Lib Magazine*. Vol. 14. no. 3/4. http://www.dlib.org/dlib/march08/marshall/03marshall-pt1.html.

[Masaaki1986] I Masaaki. 1986. *Kaizen: The key to Japan's competitive success*. MCGRAW-HILL/IRWIN. NY.

[McBride2006] S. McBride, R. Gilder, R. Davis, and S. Fenton. 2006. *"Data Mapping". Journal of the American Health Information Management Association*. Vol. 77. no. 2. pp. 44-52.

[McCartney2006] Scott McCartney. 2006. *"When Pilots Pass the BRBON, They Must Be in Kentucky"*. *The Wall Street Journal.* http://online.wsj.com/article/SB114291174429403797.html.

[McCartney2015] Scott McCartney. 15 July 2015. *"Technology will speed you through the airport of the future"*. *Wall Street Journal.*

[McCulloch2003] Mark Richard McCulloch. *Understanding W.G. Sebald.* 2003. UNIVERSITY OF SOUTH CAROLINA PRESS. Columbia.

[McDonough2006] J.P. McDonough. 2006. *METS: Standardized encoding for digital library objects.* *International Journal of Digital Libraries.* Vol. 6. no. 2. pp. 148–158.

[McGrath2015] Ben McGrath. 13 April 2015. *"Dream teams"*. *The New Yorker.* http://www.newyorker.com/magazine/2015/04/13/dream-teams.

[Medin1993] Douglas L. Medin, Robert L. Goldstone, and Dedre Gentner. 1993. *"Respects for Similarity"*. *Psychological Review.* Vol. 100. no. 2. pp. 254-278.

[Medin1997] Douglas L. Medin, E B Lynch, J D Coley, and Scott Atran.

1997. *"Categorization and reasoning among tree experts: do all roads lead to Rome?".* *Cognitive psychology.* Vol. 32. no. 1. pp. 49-96. http://www.ncbi.nlm.nih.gov/pubmed/9038245. doi: 10.1006/cogp.1997.0645.

[Melville1851] Herman Melville. *Moby Dick. or The Whale.* 1851. London.

RICHARD BENTLEY.

[Merryman2006] John Henry Merryman. 2006. *Imperialism, Art and Restitution.* Cambridge, UK.

CAMBRIDGE UNIVERSITY PRESS.

[Mervis1981] Carolyn B. Mervis and Eleanor Rosch. 1981. *"Categorization of natural objects".* *The Annual Review of Psychology.* Vol. 32. pp. 89-115.

[Miksa1984] Francis Miksa. 1984. *The development of classification at the Library of Congress.* *University of Illinois Graduate School of Library and Information Science.*

[Millen2006] David R. Millen, Jonathan Feinberg, and Bernard Kerr. 2006. *"Dogear: Social Bookmarking in the Enterprise".* *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*

(CHI '06). pp. 111-120.

[Miller1956] George Miller. 1956. *"The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information"*. *Psychological Review*. Vol. 63. no. 2. pp. 81-97.

[Miller1976] George A. Miller and Philip Johnson-Laird. 1976. *Language and Perception*. Cambridge, MA. BELKNAP PRESS.

[Miller1998] George A. Miller. 1998. *"Nouns in WordNet"*. *In WordNet: An Electronic Lexical Database*. Christine Feldbaum, (Ed.). Cambridge, MA. THE MIT PRESS.

[Miller2003] Joaquin Miller and Jishnu Mukerji. 2003. *MDA Guide Version 1.0.1*. *Object Management Group*. http://www.omg.org/cgi-bin/doc?omg/03-06-01.

[Mizzaro1997] Stefano Mizzaro. 1997. *"Relevance: The Whole History"*. *Journal of the American Society for Information Science*. Vol. 48. no. 9. pp. 810-832.

[Mockapetris1987] Paul Mockapetris. 1987. *RFC 1035:*

Domain Name System (DNS). http://tools.ietf.org/html/rfc1035.

[Monmonier1996] M. Monmonier. 1996. *How to Lie with Maps (2nd Edition).* Chicago.

UNIVERSITY OF CHICAGO PRESS.

[Montague2010] James Montague. 2010. *The rise and fall of fantasy sports.* http://www.cnn.com/2010/SPORT/football/01/06/fantasy.football.moneyball.sabermetrics/index.html.

[Morgan1871] Lewis Henry Morgan. 1871. *Systems of Consanguinity and Affinity of the Human Family.* Lincoln.

UNIVERSITY OF NEBRASKA PRESS.

[Morgan1997] Lewis Henry Morgan. 1997. *Systems of consanguinity and affinity of the human family. Reprint of Morgan(1871).*

SMITHSONIAN INSTITUTION.

[GMorgan1997] Gareth, F. Gregory, and C. Roach. 1997. *Images of Organization.*

JOHN WILLEY ONLINE.

DOI: 10.1111/j.1754-7121.1989.tb01360.x.

[Morstatter2013] F.J. Morstatter, H Liu, and K.M. Carley. 2013. *Is the sample good enough? Comparing data from Twitter's streaming API with Twitter's firehose.* *In Proceedings of International AAAI Conference on Weblogs and Social Media (ICWSM), July 8-10.* BostonMass.

[Morville2006] Peter Morville and Louis Rosenfeld. 2006. *Information Architecture for the World Wide Web.* Sebastopol, CA.

O'REILLY.

[Moussaid2009] M. Moussaid, S. Garnier, G. Theraulaz, and D. Helbing. 2009. *Collective information processing and pattern formation in swarms, flocks, and crowds..* *Topics in Cognitive Science.* Vol. . 1. Iss. 3. p 469-497.

[Munk2004] Nina Munk. 2004. *Fools Rush In: Steve Case, Jerry Levin, and the Unmaking of AOL Time Warner.* New York.

HARPERCOLLINS.

N

[Nelson1974] Theodor Holm Nelson. 1974. *Computer Lib.* Redmond, WA.

MICROSOFT PRESS.

[Nelson1981] Theodor H. Nelson. 1981. *Literary Machines: The Report On, and Of, Project Xanadu Concerning Word Processing, Electronic Publishing, Hypertext, Thinkertoys, Tomorrow's Intellectual Revolution, and Certain Other Topics Including Knowledge, Education and Freedom.* Sausalito, CA.

T. NELSON.

3rd edition.

[Neuhaus2008] Christoph Neuhaus and Hans-dieter Daniel. 2008. *"Data sources for performing citation analysis: An overview".* *Journal of Documentation.* Vol. 64. no. 2. pp. 193-210.

[Newell1972] Allen Newell and Herbert Alexander Simon. 1972.

Human Problem Solving. Englewood Cliffs, NJ.

PRENTICE HALL.

[Neziroglu2014] Fugen Neziroglu. 2014. *Hoarding: The Basics.*

ANXIETY AND DEPRESSION SOCIETY OF AMERICA.

http://www.adaa.org/understanding-anxiety/obsessive-compulsive-disorder-ocd/hoarding-basics.

[NISO2004] . 2004. *Understanding Metadata.* Baltimore, MD.

NISO PRESS.

http://www.niso.org/standards/resources/UnderstandingMetadata.pdf.

[Norman1988] Donald A. Norman. 1988. *The psychology of everyday things.* New York.

BASIC BOOKS.

[Norman1999] Donald A. Norman. 1999. *"Affordance, conventions, and design".* *interactions.* Vol. 6. no. 3. pp. 38-43.

[Norman2006] Donald A. Norman. 2006. *"Logic Versus Usage: The Case for*

Activity-Centered Design". *interactions*. Vol. 13. no. 6. pp. 45-ff.

[Norton2014] Steven Norton. 2014-07-10. *"Germany's 12th Man at the World Cup: Big Data"*. THE WALL STREET JOURNAL. http://blogs.wsj.com/dailyfix/2014/07/10/germanys-12th-man-at-the-world-cup-big-data/.

[Nunberg1996] Geoffrey Nunberg. 1996. *"Farewell to the Information Age"*. *In The Future of the Book*. Geoffrey Nunberg, (Ed.). Berkeley, CA. UNIVERSITY OF CALIFORNIA PRESS.

[Nunberg2009] Geoffrey Nunberg. 2009. *"Google's Book Search: A Disaster for Scholars"*. *The Chronicle of Higher Education*. August 31. http://chronicle.com/article/Googles-Book-Search-A/48245/.

[Nunberg2011] Geoffrey Nunberg. 2011. *"James Gleick's History of Information"*. *The New York Times*. 2011-03-18. http://www.nytimes.com/2011/03/20/books/review/book-review-the-information-by-james-gleick.html.

O

[OASIS2003] . 2003. *XML Common Biometric Format.* http://www.oasis-open.org/committees/download.php/3353/oasis-200305-xcbf-specification-1.1.doc.

[OASIS2006] . 2006. *Universal Business Language (UBL).* http://docs.oasis-open.org/ubl/os-UBL-2.0/UBL-2.0.html.

[Orwell1949] George Orwell. 1949. *Nineteen Eighty-Four (1984).* Secker and Warburg. London.

P

[Page1999] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. *“The PageRank Citation*

Ranking: Bringing Order to the Web". Stanford InfoLab. http://ilpubs.stanford.edu:8090/422/.

[Pancake2012] David Pancake. 4 Jan 2012. *Searching for Beethoven.*

WALL STREET JOURNAL.

http://www.wsj.com/articles/SB10001424052970203462304577138503300045824.

[Pandey2010] Upasana Pandey and Shampa Chakravarty. 2010. *"A Survey on Text Classification Techniques for E-mail Filtering". 2010 Second International Conference on Machine Learning and Computing.* pp. 32-36. doi: 10.1109/ICMLC.2010.61.

[Panizzi1841] Anthony Panizzi. 1841. *Rules for the Compilation of the Catalogue.*

[Panofsky1972] Erwin Panofsky and Gerda S. Panofsky. 1972. *Studies in Iconology: Humanistic Themes in the Art of the Renaissance.* Boulder, CO.

WESTVIEW PRESS.

[Park2009] Jung-Ran Park. 2009. *"Metadata Quality in Digital Repositories: A Survey* *of the Current State of the Art”. Cataloging & Classification Quarterly*. Vol. 47. (3-4). pp. 213-228. http://dx.doi.org/10.1080/01639370902737240. doi: 10.1080/01639370902737240.

[Pathan2008] Mukaddim Pathan, Rajkumar Buyya, and Athena Vakali. 2008. *“Content Delivery Networks: State of the Art , Insights, and Imperatives”*. pp. 3-32. New York.

SPRINGER.

[Perlman1984] G Perlman. 1984. *Natural artificial languages: low level processes. International Journal of Man-Machine Studies*. 20. 4.

[Pirolli2007] Peter L. T. Pirolli. 2007. *Information Foraging Theory: Adaptive Interaction with Information*. Oxford.

OXFORD UNIVERSITY PRESS.

[Plato370BC] Plato. 370 BC. *Phaedrus*. Ancient Greece.

[Pogue2009] David Pogue. 2009. *“Should You Worry About Data Rot?”. The New York Times*. 2009-03-26. http://pogue.blogs.nytimes.com/2009/03/26/should-you-worry-about-data-rot/.

[Pohs2013] Wendi Pohs. 2013. *"Building a taxonomy for auto-classification"*. *Bulletin of the American Societity for Information Science and Technology*. Vol. 39. no. 2. pp. 34-38. doi: 10.1002/bult.2013.1720390210.

[Polaine2013] A. Polaine, L Løvlie, and B. Reason. 2013. *Service design: from insight to implementation*. Brooklyn, NY.

ROSENFELD MEDIA.

[Poole2010] Erika Shehan Poole and Jonathan Grudin. 2010. *"A taxonomy of Wiki genres in enterprise settings"*. *Proceedings of the 6th International Symposium on Wikis and Open Collaboration – WikiSym '10*. http://portal.acm.org/citation.cfm?doid=1832772.1832792. doi: 10.1145/1832772.1832792.

[Pope2011] Julia T. Pope and Robert P. Holley. 2011. *"Google Book Search and Metadata"*. *Cataloging & Classification Quarterly*. Vol. 49. no. 1. pp. 1-13.

[Pralahad1990] C. K. Pralahad and Gary Hamel. 1990. *"The Core Competence of the Corporation"*.

Harvard Business Review. http://hbr.org/1990/05/the-core-competence-of-the-corporation/ar/1.

[Prats2008] Mario Prats, Ester Martínez, Pedro J. Sanz, and Angel P. Pobil. 2008. *“The UJI librarian robot”*. *Intelligent Service Robotics*. Vol. 1. no. 4. pp. 321-335. http://www.springerlink.com/content/15164724vxp07704/.

[PREMIS2011] . 2008. *Data Dictionary for Preservation Metadata: PREMIS version 2.0*. Washington, DC.

LIBRARY OF CONGRESS.

http://www.loc.gov/standards/premis/.

[Provost2013] Foster Provost and Tom Fawcett. 2013. *Data Science for Business*. *What you need to know about data mining and data-analytic thinking*.

O’REILLY MEDIA.

[Proctor2011] Nancy Proctor. 2011. *“The Google Art Project: A New Generation of Museums on the Web?”*. *Curator: The Museum Journal*. Vol. 54. no. 2. pp. 215-221. doi: 10.1111/j.2151-6952.2011.00083.x.

Q

[Queenan2011] Joe Queenan. 2011. *“Wotan, Your Double-Skim Latte Is Ready”. The Wall Street Journal.* 2011-09-24. http://online.wsj.com/article/SB10001424053111904106704576582834147448392.html.

R

[Rahm2000] Erhard Rahm and Hong Hai Do. 2000. *“Data cleaning: Problems and current approaches”. IEEE Data Eng. Bull.* 23. no. 4.

[Rahm2006] Erhard Rahm and Philip A. Bernstein. 2006. *“An Online Bibliography on Schema Evolution”. ACM SIGMOD Record.* Vol. 35. no. 4. pp. 30-31.

[Ranganathan1967] Shiyali Ramamrita Ranganathan. 1967.

Prologmena to Library Classification. Delhi.

ASIA PUBLISHING HOUSE.

[RDFWorkingGroup2004] . 2004. *Resource Description Framework*. *Recommendation of the World Wide Web Consortium (W3C)*. http://www.w3.org/RDF/.

[Reaney1997] P. H. Reaney and R. M. Wilson. 1997. *A Dictionary of English Surnames*. Oxford.

OXFORD UNIVERSITY PRESS.

[Regazzoni2010] Carlo S. Regazzoni, Andrea Cavallaro, Ying Wu, and Janusz Konrad. 2010. *"Video Analytics for Surveillance: Theory and Practice"*. *IEEE Signal Processing*. Vol. 27. no. 5. pp. 16–17.

[Register2012] Renée Register and Thad McIlroy. 2012. *The Metadata Handbook: A Book Publisher's Guide to Creating and Distributing Metadata for Print and Ebooks*. Columbus, OH.

DATACURATE.

[Rehder2004] Bob Rehder and Reid Hastie. 2004. *"Category coherence and category-based property induction"*. *Cognition*. Vol. 91. no. 2. pp. 113-153. http://linkinghub.elsevier.com/retrieve/pii/S0010027703001677. doi: 10.1016/S0010-0277(03)00167-7.

[Renear2003] Allen Renear and David Dubin. 2003. *"Towards Identity Conditions for Digital Documents"*. *International Conference on Dublin Core and Metadata Applications*. pp. 181-189. http://dcpapers.dublincore.org/index.php/pubs/article/view/746.

[Resmini2011a] Andrea Resmini and Luca Rosati. 2011. *A Brief History of Information Architecture*. *Journal of Information Architecture*. 3 (2). 33-46.

[Resmini2011b] Andrea Resmini and Luca Rosati. 2011. *Pervasive Information Architecture: Designing Cross-Channel User Experiences*. ELSEVIER.

[Resnick2001] P. Resnick. 2001. *“Internet Message Format”. The Internet Society.*

[Reynolds2014] Gretchen Reynolds. 2014-07-16. *“Train Like a German Soccer Star”.*

THE NEW YORK TIMES.

http://nyti.ms/1rfy8rd.

[Rips2012] Lance J. Rips, Edward E. Smith, and Douglas L. Medin. 2012. *“Concepts and Categories: Memory, Meaning, and Metaphysics”. In The Oxford Handbook of Thinking and Reasoning.* Keith J. Holyoak and Robert G. Morrison. pp. 177-209. New York.

OXFORD UNIVERSITY PRESS, USA.

[Robbins2016] Lenn Robbins. 27 April 2016. *“A Data Scientist Dissects the 2016 NFL Draft”. Wall Street Journal.*

http://www.wsj.com/articles/a-data-scientist-dissects-the-2016-nfl-draft-1461793878

.

[Robertson2015] Brian J. Robertson. 2015. *Holacracy: The New Management System for a Rapidly Changing World.*

HENRY HOLT AND COMPANY.

[Robertson2005] Stephen Robertson. 2005. *"How Okapi came to TREC"*. *In TREC: Experiment and Evaluation in Information Retrieval*. Ellen M. Voorhees and Donna K. Harman, (Eds.). 287-299. Cambridge, MA.

THE MIT PRESS.

[Robinson2008] David Robinson, Harlan Yu, William P. Zeller, and Edward W. Felten. 2008. *"Government Data and the Invisible Hand"*. *Yale Journal of Law and Technology*. Vol. 11. pp. 160-176. http://papers.ssrn.com/abstract=1138083.

[Rogers2008] Timothy T. Rogers and James L. Mcclelland. 2008. *"Précis of Semantic Cognition: A Parallel Distributed Processing Approach"*. *Behavioral and Brain Sciences*. Vol. 31. no. 6. pp. 689-749.

[Rosch1975] Eleanor Rosch. 1975. *"Cognitive representations of semantic categories"*. *Journal of Experimental Psychology: General*. Vol. 104. no. 3. pp. 192-233. http://psycnet.apa.org/journals/xge/104/3/192/. doi: 10.1037/0096-3445.104.3.192.

[Rosch1999] Eleanor Rosch. 1999. *"Principles of Categorization"*. *In Concepts: Core Readings*

(Margolis 1999). Eric Margolis and Stephen Laurence. pp. 189-206. Cambridge, MA.

THE MIT PRESS.

[Rose2016] Tom Rose. 2016. *The End of Average. How We Succeed in a World of Sameness.*

HARPER COLLINS.

[Rosen2012] Jeffrey Rosen. 2012. *"Who Do Online Advertisers Think You Are?". The New York Times Magazine.* 2012-12-02. http://www.nytimes.com/2012/12/02/magazine/who-do-online-advertisers-think-you-are.html.

[Rosenthal2004] Arnon Rosenthal, Len Seligman, and Scott Renner. 2004. *"From semantic integration to semantics management: case studies and a way forward". ACM SIGMOD Record.* Vol. 33. no. 4. pp. 44-50. doi: 10.1145/1041410.1041418.

[Rothenberg1999] Jeff Rothenberg. 1999. *"Ensuring the Longevity of Digital Information". RAND.*

[Rothfarb2007] Robert J. Rothfarb and Paul Doherty. 2007. *"Creating Museum Content and Community in Second Life". Museums and the Web 2007: International Conference for Culture* *and Heritage On-line.* http://www.museumsandtheweb.com/mw2007/papers/rothfarb/rothfarb.html.

[Rowley2007] Jennifer E. Rowley. 2007. *"The wisdom hierarchy: representations of the DIKW hierarchy". Journal of Information Science.* 33. no. 2. pp. 163-180.

[Rowling1997-2007] J.K. Rowling. 1997-2007. *Harry Potter Series.* United Kingdom. BLOOMSBURY PUBLISHING.

[Rowling2013] J.K. Rowling, writing as Robert Galbraith. 2013. *The Cuckoo's Calling.* United Kingdom. SPHERE BOOKS.

[Rubinsky1997] Yuri Rubinsky and Murray Maloney. 1997. *SGML on the Web: Small Steps Beyond HTML.* Upper Saddle River, NJ. PRENTICE HALL.

S

[Sag2012] Matthew Sag. 2012. *“Orphan Works as Grist for the Data Mill”*. *SSRN Electronic Journal.* http://papers.ssrn.com/abstract=2038889.

[Salton1975] Gerard Salton, Anita Wong, and Chung-Shu Yang. 1975. *“A Vector Space Model for Automatic Indexing”*. *Communications of the ACM.* Vol. 18. no. 11. pp. 613-320.

[Samuelson2009] Pamela Samuelson. 2009. *Google Books Is Not a Library*. *Huffington Post.* http://www.huffingtonpost.com/pamela-samuelson/google-books-is-not-a-lib_b_317518.html.

[Samuelson2010] Pamela Samuelson. 2010. *“Google Book Search and the Future of Books in Cyberspace”*. *Miinnesota Law Review.* Vol. 94. no. 5. pp. 1308-1374.

[Samuelson2011] Pamela Samuelson. 2011. *“The Google Book Settlement as Copyright Reform”*. *Wisconsin Law Review.* Vol. 11. no. 2. pp.

479-562.
http://papers.ssrn.com/abstract=1683589.

[Sandler2006] Mark Sandler. 2006. *"Collection Development in the Age Day of Google". Library Resources & Technical Services.* Vol. 50. no. 4. pp. 239.
http://connection.ebscohost.com/c/articles/22421812/collection-development-age-day-google.

[Saracevic1975] Tefko Saracevic. 1975. *"Relevance: A Review of the Literature and a Framework for Thinking on the Notion in Information Science". Journal of the American Society for Information Science.* Vol. 26. no. 6. pp. 321-343. doi: 10.1002/asi.

[Satija2001] Mohinder P. Satija. 2001. *"Relationships in Raganathan's Colon Classification". In Relationships in the Organization of Knowledge.* Carol A. Bean and Rebecca Green, (Eds.). 199-210. Norwell, MA.

KLUWER.

[Savodnik2011] Peter Savodnik. 2011. *"Knut, the $140 Million Polar Bear". Bloomberg*

BusinessWeek. May 26. http://www.petersavodnik.com/articles/knut-the-140-million-polar-bear/.

[Sayre2005] Robert Sayre. 2005. *“Atom: The Standard in Syndication”*. *IEEE Internet Computing*. Vol. 9. no. 4. pp. 71-78.

[Schamber1990] Linda Schamber, Michael B. Eisenberg, and Michael S. Nilan. 1990. *“A re-examination of relevance: toward a dynamic, situational definition”*. *Information Processing & Management*. Vol. 26. no. 6. pp. 755-776. http://linkinghub.elsevier.com/retrieve/pii/030645739090050C. doi: 10.1016/0306-4573(90)90050-C.

[Schatz1994] Bruce R. Schatz and Joseph B. Hardin. 1994. *“NCSA Mosaic and the World Wide Web: Global Hypermedia Protocols for the Internet”*. *Science*. Vol. 265. no. 5174. pp. 895-901. doi: 10.1126/science.265.5174.895.

[Schmandt-Besserat1997] Denise Schmandt-Besserat. 1997. *How Writing Came About*. Austin.

UNIVERSITY OF TEXAS PRESS.

[Schmidt2009] Desmond Schmidt. 2009. *“Merging Multi-Version Texts: A General*

Solution to the Overlap Problem". Proceedings of Balisage: The Markup Conference 2009. http://www.balisage.net/Proceedings/vol3/print/Schmidt01/BalisageVol3-Schmidt01.html.

[Schmitz2008] Patrick L. Schmitz and Michael T. Black. 2008. *"The Delphi Toolkit: Enabling Semantic Search for Museum Collections". Museums and the Web 2008: International Conference for Culture and Heritage On-line.*

[Scholl2007] Hans Jochen Scholl and Ralf Klischewski. 2007. *"International Journal of Public E-Government Integration and Interoperability: Framing the Research Agenda". International Journal of Public Administration.* Vol. 30. (8-9). pp. 889-920. doi: 10.1080/01900690701402668.

[Schreibman2005] Susan Schreibman, Ray Siemens, and John Unsworth. 2005. *A Companion to Digital Humanities.* Hoboken, NJ. WILEY-BLACKWELL.

[Schwartz2005] Barry Schwartz. 2005. *The Paradox of Choice: Why More Is Less.* New York.

HARPER PERENNIAL.

[Sebald1995] W.G. Sebald. 1995. *The Rings of Saturn. (English Ed.). (Die Ringe des Saturn. Eine englische Wallfahrt.).*

HARVILL. LONDON.

[SeeLi2006] Jiexun SeeLi, Rong Zheng, and Hsinchun Chen. 2006. *“From fingerprint to writeprint”. Communications of the ACM.* 49. no, 4. pp. 76-82. dl.acm.org/citation.cfm?id=1121951

.

[Sen2004] Arun Sen. 2004. *“Metadata management: past, present and future”. Decision Support Systems.* Vol. 37. no. 1. pp. 151-173. http://dx.doi.org/10.1016/S0167-9236(02)00208-7. doi: 10.1016/S0167-9236(02)00208-7.

[Sen2006] Shilad Sen, Shyong K. Lam, Al Mamunur Rashid, Dan Cosley, Dan Frankowski, Jeremy Osterhouse, F. Maxwell Harper, and John Riedl. 2006. *“tagging, communities, vocabulary, evolution”. CSCW ’06: Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work.* pp. 181-190.

[Settles2012] Burr Settles. 2012. *“Active learning”.*

Synthesis Lectures on Artificial Intelligence and Machine Learning. 7. no. 1. p 1-114.

[Shadbolt2006] Nigel Shadbolt, Tim Berners-Lee, and Wendy Hall. 2006. *“The Semantic Web Revisited”*. *IEEE Intelligent Systems*. Vol. 21. no. 3. pp. 96-101. doi: 10.1109/MIS.2006.62.

[Shah2006] Rajiv C. Shah and Jay P. Kesan. 2006. *“Open Standards and the Role of Politics”*. *The Proceedings of the 8th Annual International Digital Government Research Conference*.

[Shamurin1955] Evgenii Ivanovich Shamurin. 1955. *Ocherki po istorii bibliotechno-bibliograficheskoi klassifikatsii [Notes on the history of library and bibliographic classification].*. Moskva [Moscow].

VSESOIUZNAIA KNIZHNAIA PALATA [ALL-UNION BOOK CHAMBER].

[Shannon1948] Claude E. Shannon. July 1948. *“A Mathematical Theory of Communication”*. *Bell System Technical Journal*. 27. no. 3. pp 379–423.

[Shakespeare1623] William Shakespeare. *The Tragedie of Macbeth*. *Folio*. 1623. London.

[Shapiro1998] Carl Shapiro and Hal R. Varian. 1998. *Information rules: a strategic guide to the network economy*. Cambridge, MA.

HARVARD BUSINESS REVIEW PRESS.

[Shiner2007] Larry Shiner. 2007. *"Architecture vs. Art: The Aesthetics of Art Museum Design"*. *Contemporary Aesthetics*. 5.

[Shirky2005] Clay Shirky. 2005. *Ontology Is Overrated: Categories, Links, and Tags*. http://www.shirky.com/writings/ontology_overrated.html.

[Siegel2013] E. Siegal. 2013. *Predictive analytics: The power to predict who will click, buy, lie, or die.*. Hoboken, NJ.

JOHN WILEY & SONS.

[Silman1998] Roberta Silman. 1998. *"In the Company of Ghosts"*. *The New York Times*. 1998-07-26.

http://www.nytimes.com/books/98/07/26/reviews/980726.26silmant.html.

[Silver2012] Nate Silver. 2012. *The Signal and the Noise. Why So Many Predictions Fail—but Some Don't.*

PENGUIN PRESS.

[Silverman2013] Rachel Silverman. 6 August 2013. *Some Tech Firms Ask: Who Needs Managers?.*

WALL STREET JOURNAL.

http://on.wsj.com/1zcS3tb.

[Silverston2000] Len Silverston. 2000. *The Data Model Resource Book, Vol. 2: A Library of Data Models for Specific Industries.* Hoboken, NJ.

JOHN WILEY & SONS.

[Simon1982] Herbert Alexander Simon. 1982. *Models of bounded rationality: Empirically grounded economic reason.* Vol. 3.

MIT PRESS.

[Simon1996] Herbert Alexander Simon. 1996. *The Sciences of the Artificial.* Cambridge, MA.

THE MIT PRESS.

[Simon1997] Herbert Alexander Simon. 1997. *Administrative Behavior*. New York.

FREE PRESS.

[Simon2010] Nina Simon. 2010. *The Participatory Museum*. Santa Cruz, CA.

MUSEUM 2.0.

http://www.participatorymuseum.org/read/.

[Simpson1989] John Simpson and Edmund Weiner. 1989. *The Oxford English Dictionary*. Oxford.

OXFORD UNIVERSITY PRESS.

[Sinclair2007] James Sinclair and Michael Cardew-Hall. 2007. *"The folksonomy tag cloud: when is it useful?"*. *Journal of Information Science*. Vol. 34. no. 1. pp. 15-29. doi: 10.1177/0165551506078083.

[Smiraglia1994] Richard P. Smiraglia. 1994. *"Derivative Bibliographic Relationships: Linkages in the Bibliographic Universe"*. *Navigating the Networks: Proceedings of the ASIS Mid-Year Meeting, Portland, Oregon, May 21-25, 1994*. pp. 115-135.

[Smiraglia1999] Richard P. Smiraglia and Gregory H. Leazer. 1999. *“Derivative Bibliographic Relationships: The Work Relationship in a Global Bibliographic Database”. Journal of the American Society for Information Science.* Vol. 50. no. 6. pp. 493-504.

[Smith1776] Adam Smith. 1776. *Wealth of Nations.* London.

W. STRAHAN AND T. CADELL.

[Smith1981] Edward E. Smith and Douglas L. Medin. 1981. *Categories and Concepts.* Cambridge, MA.

HARVARD UNIVERSITY PRESS.

[Smith2003] Linda B. Smith and Esther Thelen. 2003. *“Development as a dynamic system”. Trends in Cognitive Sciences.* Vol. 7. no. 8. pp. 343-348. http://linkinghub.elsevier.com/retrieve/pii/S1364661303001566. doi: 10.1016/S1364-6613(03)00156-6.

[Spence1985] Johnathan D. Spence. 3 September 1985. *The Memory Palace of Matteo Ricci.*

PENGUIN BOOKS.

[Spencer2009] D. Spencer. 2009. *Card Sorting.* Brooklyn, NY.

ROSENFELD MEDIA.

[Spiteri1998] Louis Spiteri. 1998. *“A simplified model of facet analysis”*. *Canadian Journal of Information and Library Science*. Vol. 23. (April-July). pp. 1-30.

[Srinivasan2009] Ramesh Srinivasan, Robin Boast, Jonathan Furner, and Katherine M. Becvar. 2009. *“Digital museums and diverse cultural knowledges: Moving past the traditional catalog”*. *The Information Society*. Vol. 25. no. 4. pp. 265-278.

[Stewart1997] Thomas A. Stewart. 1997. *Intellectual Capital: The New Wealth of Organizations*. New York.

CROWN BUSINESS.

[Storey1993] Veda C. Storey. 1993. *“Understanding Semantic Relationships”*. *VDLB Journal*. Vol. 2. pp. 455-488.

[Strout1956] Ruth French Strout. 1956. *“The development of the catalog and cataloging codes”*. *The Library Quarterly*. Vol. 26. no. 4. pp. 254-275.

[Suehle2012] Ruth Suehle. 2012. *“The Periodic Tables* *of Everything but Elements”. Wired.* March 1. http://www.wired.com/geekmom/2012/03/the-periodic-tables-of-everything-but-elements/.

[Suits1967] Bernard Suits. 1967. *“What is a Game?”. Philosophy of Science.* pp. 148-156.

[Svenonius2000] Elaine Svenonius. 2000. *The Intellectual Foundation of Information Organization.* Cambridge, MA.

THE MIT PRESS.

T

[Tagliabue2012] John Tagliabue. 2012. *“Swiss Cows Send Texts to Announce They’re in Heat”. The New York Times.* 2012-10-01. http://www.nytimes.com/2012/10/02/world/europe/device-sends-message-to-swiss-farmer-when-cow-is-in-heat.html.

[Talukder2016] Hisham Talukder. 2016. *“Preventing in--game injuries for NBA players”. MIT Sports Analytics Conference.* http://www.sloansportsconference.com/wp-content/uploads/2016/02/1590-Preventing-in-game-injuries-for-NBA-players.pdf
.

[Tauberer2014] Joshua Tauberer. 2014. *Open Government Data: The book.* https://opengovdata.io/2014/history-the-movement/
.

[Taylor2009] Arlene G. Taylor and Daniel N. Joudrey. 2009. *The Organization of Information.* Santa Barbara, CA.

LIBRARIES UNLIMITED.

[Taylor1914] F.W. Taylor. 1914. *The principles of scientific management.*

HARPER.

[Taylor2006] Hugh Taylor. 2006. *The Joy of SOX: Why Sarbanes-Oxley and Service-Oriented Architecture May*

Be the Best Thing That Ever Happened to You. Hoboken, NJ.

WILEY.

[Taylor2007] James Taylor and Neil Raden. 2007. *Smart (Enough) Systems: How to Deliver Competitive Advantage by Automating Hidden Decisions.* Upper Saddle River, NJ.

PRENTICE HALL.

[Taylor2010] Maureen A. Taylor. 2010. *Preserving Your Family Photographs.* N.p.

PICTURE PERFECT PRESS.

[Teece1998] David J. Teece. 1998. *"Capturing Value from Knowledge Assets: The New Economy, Markets for Know-how, and Intangible Assets".* *California Management Review.* Vol. 40. no. 3. pp. 55-79.

[Tenenbaum2000] J.B. Tenenbaum. 2000. *"Rules and similarity in concept learning".* *Advances in neural information processing systems.* 12. pp. 59-65.

[Tenenbaum2001] J.B. Tenenbaum and T.L. Griffiths. 2001. *"Generalization, similarity, and Bayesian inference". Behavioral and brain sciences.* Vol. 24. no. 04. pp. 629-640.

[Teper2005] Thomas H. Teper. 2005. *"Current and emerging challenges for the future of library and archival preservation". Library Resources & Technical Services.* Vol. 49. no. 1. pp. 32-39.

[Thaler2008] Richard H. Thaler and Cass R. Sunstein. 8 April 2008. *Nudge: Improving Decisions about Health, Wealth and Happiness.*

YALE UNIVERSITY PRESS.

[Tidwell2008] Doug Tidwell. 2008. *XSLT: Mastering XML Transformations.* Sebastopol, CA.

O'REILLY.

[Tidwell2010] Jennifer Tidwell. 2010. *Designing Interfaces.* (2nd Edition). Sebastopol, CA.

O'REILLY.

[Tillett1991] Barbara B. Tillett. 1991. *"A Taxonomy of*

Bibliographic Relationships". Library Resources & Technical Services. Vol. 35. no. 2. pp. 150-158.

[Tillett1992] Barbara B. Tillett. 1992. *"Bibliographic relationships: An empirical study of the LC machine-readable records". Library Resources & Technical Services.* Vol. 36. no. 2. pp. 162-188.

[Tillett2001] Barbara B. Tillett. 2001. *"Bibliographic Relationships". In Relationships in the Organization of Knowledge.* Carol A. Bean and Rebecca Green, (Eds.). 19-36. Norwell, MA.

KLUWER.

[Tillett2003] Barbara Tillett. 2003. *"What is FRBR? A Conceptual Model for the Bibliographic Universe". Technicalities.* Vol. 25. no. 5.

[Tillett2005] Barbara B. Tillett. 2005. *"FRBR and Cataloging for the Future". Cataloging & Classification Quarterly.* Vol. 39. no. 3/4. pp. 197-205.

[Toma2008] Catalina L. Toma, Jeffrey T. Hancock, and Nicole B. Ellison. 2008. *"Separating fact from fiction: An examination of deceptive* *self-presentation in online dating profiles"*. *Personality and Social Psychology Bulletin*. 34.8. pp 1023-1036.

[Trant2009] Jennifer Trant. 2009. *"Emerging convergence? Thoughts on museums, archives, libraries, and professional training"*. *Museum Management and Curatorship*. Vol. 24. no. 4. pp. 1-24. doi: 10.1080/09647770903314738.

[Travers1969] Jeffrey Travers and Stanley Milgram. 1969. *"An experimental study of the small world problem"*. *Sociometry*. Vol. 32. no. 4. pp. 425-443.

[Trofimov2015] Y Trofimov. 10 April 2015. *"Would new borders mean less conflict in the Middle East?"*. *Wall Street Journal*. http://on.wsj.com/1DsapZP.

[Turban2010] Efraim Turban, Ramesh Sharda, and Dursun Delen. 2010. *Decision Support and Business Intelligence Systems*. Upper Saddle River, NJ. PRENTICE HALL.

[Tufte1983] Edward R. Tufte. 1983. *The Visual Display of Quantitative Information.* Cheshire, CT.

GRAPHICS PRESS.

[Turnbull2009] Giles Turnbull. 2009. *"A Common Nomenclature for Lego Families".* *The Morning News.* 2009. http://www.themorningnews.org/article/a-common-nomenclature-for-lego-families.

[Turney2008] Peter D. Turney. 2008. *"The latent relation mapping engine: Algorithm and experiments.".* *Journal of Artificial Intelligence Research.* pp. 615-655. .

[Tversky1974] Amos Tversky and Daniel Kahneman. 1974. *"Judgment under uncertainty: Heuristics and biases".* *Science.* Vol. . 185 . no. 4157.

[Tyson2011] *The Pluto Files.* 2011. Neil deGrasse Tyson and Terri Randall.

BOSTON, MA: WGBH.

U

[Underhill2008] Paco Underhill. 2008. *Why We Buy: The Science of Shopping – Updated and Revised for the Internet, the Global Consumer, and Beyond.* New York.

SIMON & SCHUSTER.

V

[vanderVlist2007] Eric van der Vlist. 2007. *Schematron.* Sebastopol, CA.

O'REILLY.

[VanderWal2007] Thomas Vander Wal. 2007. *Folksonomy. vanderwal.net.* http://vanderwal.net/folksonomy.html.

[VandeSompel2010] Herbert VandeSompel, Michael Nelson, and Robert

Sanderson. 2010. *HTTP framework for time-based access to resource states—Memento. IETF Tools.* http://tools.ietf.org/html/draft-vandesompel-memento-00.

[VanDuyne2006] Douglas K. VanDuyne, J.A. Landay, and J.I. Hong. 2006. *The Design of Sites: Patterns for Creating Winning Web Sites.* (2nd Edition). Upper Saddle River, NJ.

PRENTICE HALL.

[Vargo2004] Stephen Vargo and Robert F. Lusch. 2004. *“Evolving to a new dominant logic for marketing.”. Journal of marketing.* 1-17.

[Venturi1972] R. Venturi, D.S. Brown, and S. Izenour. 1972. *Learning from Las Vegas.* (Vol. 102). Cambridge, MA.

MIT PRESS.

[Viswanadham2002] N. Viswanadham. 2002. *“The past, present, and future of supply-chain automation”. IEEE Robotics & Automation Magazine.* Vol. 9. no. 2. pp. 48-56. http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=1019490&contentType=Journals+&+Magazines.

[vonAhn2004] Luis von Ahn and Laura Dabbish. 2004. *"Labeling images with a computer game". Proceedings of the 2004 Conference on Human Factors in Computing Systems.* pp. 319-326.

[vonAhn2008] Luis von Ahn and Laura Dabbish. 2008. *"Designing games with a purpose". Communications of the ACM.* Vol. 51. no. 8. pp. 57-67.

[vonRiegen2006] Claus von Riegen. 2006. *How Standards Address Interoperability Needs: An Industry View. OASIS Symposium, May 10, 2006, San Francisco.*

W

[Wagemans2012] Johan Wagemans, James H. Elder, Michael Kubovy, Stephen E. Palmer, Mary A. Peterson, Manish Singh, and Rüdiger von der Heydt. 2012. *"A century of Gestalt psychology in visual perception: I.*

Perceptual grouping and figure–ground organization.”. Psychological bulletin. 138. no. 6. p 1172.

[Wakabayashi2011] Daisuke Wakabayashi. 2011. *“Japanese Farms Look to the ‘Cloud’”. The Wall Street Journal.* 2011. http://online.wsj.com/article/SB10001424052748704029704576087910899748444.html.

[Walker2009] Rob Walker. 2009. *“The Song Decoders at Pandora”. The New York Times.* 2009-10-18. http://www.nytimes.com/2009/10/18/magazine/18Pandora-t.html.

[Wallach1998] Alan Wallach. 1998. *Exhibiting Contradiction.* Amherst, MA. UNIVERSITY OF MASSACHUSETTS PRESS.

[Walsh2010] Norman Walsh. 2010. *DocBook 5: The Definitive Guide.* Sebastopol, CA. O’REILLY.

[Want2006] Roy Want. 2006. *“An Introduction to RFID Technology”. Pervasive Computing.* Vol. 5. no. 1. pp. 25-33.

[Watson2007] Hugh J. Watson and Barbara H. Wixom. 2007. *“The

Current State of Business Intelligence". *Computer*. Vol. 40. no. 9. pp. 96-99. doi: 10.1109/MC.2007.331.

[Watts2004] Duncan J. Watts. 2004. *"The 'New' Science of Networks"*. *Annual Review of Sociology*. Vol. 30. no. 1. pp. 243-270. doi: 10.1146/annurev.soc.30.020404.104342.

[Weick1995] Karl E. Weick. 1985. *"Sensemaking in organizations"*. *Sage*. Volume 3.

[Weick2005] Karl E. Weick, Kathleen M. Sutcliffe, and David Obstfeld. 2005. *"Organizing and the process of sensemaking"*. *Organization science*. 16. no. 4. pp. 409-42.

[Weill2004] Peter Weill and Jeanne Ross. 2004. *IT Governance: How Top Performers Manage IT Decision Rights for Superior Results*. Cambridge, MA.

HARVARD BUSINESS REVIEW PRESS.

[Weinberger1985] Jerry Weinberger. 1985. *Science, Faith, and Politics: Francis Bacon and the Utopian Roots of the Modern Age;*

A Commentary on Bacon's Advancement of Learning. Ithaca, NY.

CORNELL UNIVERSITY PRESS.

[Weinberger2007] David Weinberger. 2007. *Everything Is Miscellaneous: The Power of the New Digital Disorder*. New York.

TIMES BOOKS.

[Weinreich2001] Harald Weinreich, Hartmut Obendorf, and Winfried Lamersdorf. 2001. *"The Look of the Link—Concepts for the User Interface of Extended Hyperlinks"*. *Proceedings of the 12th ACM Conference on Hypertext and Hypermedia (HYPERTEXT '01)*. pp. 19-28.

[Wheatley2004] Malcolm Wheatley. 2004. *"Operation Clean Data"*. *CIO*. September 10. http://www.cio.com.au/article/166533/operation_clean_data/.

[Whistler1890] James Abbott Whistler. 1890.

PROJECT GUTENBERG.

http://www.gutenberg.org/ebooks/24650.

[Wilbert2006] Caroline Wilbert. 2006. *How Wal-Mart Works*. *HowStuffWorks.com*. http://money.howstuffworks.com/wal-mart.htm.

[Wilde2002] Erik Wilde and David Lowe. 2002. *XPath, XLink, XPointer, and XML: A Practical Guide to Web Hyperlinking and Transclusion.* Boston, MA.

ADDISON WESLEY.

[Wilde2007] Erik Wilde and Philippe Cattin. 2007. *“Presenting in HTML”.* *DocEng ’07: Proceedings of the 2007 ACM Symposium on Document Engineering.*

[Wilde2008a] Erik Wilde. 2008. *The Plain Web.* http://dret.net/netdret/docs/wilde-wsw2008/.

[Wilde2008b] Erik Wilde and Robert J. Glushko. 2008. *“Document design matters”.* *Communications of the ACM.* Vol. 51. no. 10. pp. 43-49. http://portal.acm.org/citation.cfm?doid=1400181.1400195. doi: 10.1145/1400181.1400195.

[Wilde2011] Erik Wilde and Cesare Pautasso. 2011. *REST: From Research to Practice.* New York.

SPRINGER.

[Wilkins1668] John Wilkins. 1668. *“Essay towards a*

Real Character and a Philosophical Language”.

THE ROYAL SOCIETY.

[Williams2012] Robin Williams. 2012. *The Non-Designer’s Design Book.* San Francisco, CA.

PEACHPIT PRESS.

[Williamson1975] Oliver E. Williamson. 1975. *“Markets and hierarchies: analysis and antitrust implications: a study in the economics of internal organization”.* *University of Illinois at Urbana-Champaign’s Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship.* http://ssrn.com/abstract=1496220.

[Williamson1998] Oliver E. Williamson. 1998. *The Economic Institutions of Capitalism.* New York.

FREE PRESS.

[Wilkinson2014] Alec Wilkinson. 2014 May 19. *A Voice from the Past.* *How a physicist resurrected the earliest recordings.* New York, NY.

NEW YORKER MAGAZINE.

http://www.newyorker.com/magazine/2014/05/19/a-voice-from-the-past.

[Wilson1968] Patrick Wilson. 1968. *Two Kinds of Power: An Essay on Bibliographical Control*. Berkeley.

UNIVERSITY OF CALIFORNIA PRESS.

[Winkler2010] Stefan Winkler and Jens von Pilgrim. 2010. *"A survey of traceability in requirements engineering and model-driven development". Software and Systems Modeling*. 9. no. 4. pp. 529-565.

[Winner1980] Langdon Winner. 1980. *"Do artifacts have politics?". Daedelus*.

[Wittgenstein2002] Ludwig Wittgenstein. 2002. *"Philosophical Investigations, Sections 65-78". In Foundations of Cognitive Psychology: Core Readings*. Daniel J. Levitin. pp. 271-276. Cambridge, MA.

THE MIT PRESS.

[Winston1987] Morton E. Winston, Roger Chaffin, and Douglas Herrmann. 1987. *"A taxonomy of part-whole relations". Cognitive Science*. 11. no. 4. 417-444.

[Wright2010] Alex Wright. 2010. *“Managing Scientific Inquiry in a Laboratory the Size of the Web”*. *The New York Times*. 2010-12-28. http://www.nytimes.com/2010/12/28/science/28citizen.html.

[Wu2012] Michael Wu. 2012. *The Science of Social: Beyond Hype, Likes, and Followers*. Emeryville, CA.

LITHIUM TECHNOLOGIES.

[Wynholds2011] Laura Wynholds. 2011. *“Linking to Scientific Data: Identity Problems of Unruly and Poorly Bounded Digital Objects”*. *The International Journal of Digital Curation*. Vol. 6. no. 1. pp. 214-225.

X

Y

[Yang2010] Xin-She Yang. 2010. *Nature-Inspired Optimization Algorithms.*

LUNIVER PRESS.

[Yanovsky2014] David Yanovsky. 10 June 2014. *Here are the 32 Countries Google Maps Won't Draw Borders Around.*

QUARTZ.

http://qz.com/218675/here-are-the-32-countries-google-maps-wont-draw-borders-around/.

[Yau2011] Nathan Yau. 2011. *Visualize This: The FlowingData Guide to Design, Visualization, and Statistics.* Hoboken, NJ.

WILEY.

[Yee2008] Raymond Yee. 2008. *Pro Web 2.0 Mashups: Remixing Data and Web Services.* New York.

APRESS.

[Yee2009] Martha M. Yee. 2009. *"Can Bibliographic Data Be Put Directly Onto the Semantic Web?". UC Los Angeles, California Digital Library eScholarship.* http://escholarship.org/uc/item/91b1830k.

Z

[Zeithami2001] Valarie A. Zeithami, Roland T. Rust, and Katherine N. Lemon. 2001. *"Customer Pyramid: Creating and Serving Profitable Customers". California Management Review.* Vol. 43. no. 4. pp. 118-142.

[Zhou2003] Rongron Zhou and Dilip Soman. 2003. *"Looking Back : Exploring the Psychology of*

Queuing and the Effect of the Number of People Behind". *Journal of Consumer Research*. Vol. 29. no. 4. pp. 517-530.

[Zhu2014] H. Zhu, S.E. Madnick, Y.W. Lee, and R.Y. Wang. 2014. "*Data and Information Quality Research: Its Evolution and Future*".

Glossary

MURRAY MALONEY

An Appendix to The Discipline of Organizing

Note

The glossary presents an alphabetical listing of entries, each with a term and a corresponding meaning. As much as possible and wherever practical, the contents of the glossary definitions are transcluded from the chapters and hypertext links offer the gentle reader with a means to go back to the source so that it can be considered in the author's intended context.

In the case of abbreviations, the meaning transcludes the expanded form of the abbreviation. Where abbreviations relate to formal organizations or standards, we provide a reference URI to encourage discovery.—MM

G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z

A

AACR2

Anglo-American Cataloguing Rules (AACR2)

(http://www.aacr2.org/)

AAP

Association of American Publishers (AAP)

(http://www.publishers.org/)

AAT

Art and Architecture Thesaurus (AAT)

(http://www.getty.edu/research/tools/vocabularies/aat/)

aboutness

"Subject matter" organization involves the use of a classification system that provides categories and descriptive terms for indicating what a resource is about. Because they use *aboutness* properties that are not directly perceived, methods for assigning subject classifications are intellectually-intensive and in many cases require rigorous training to be performed consistently and appropriately. (From "Organizing Resources".)

absolute synonyms

The strictest definition is that *synonyms* "are words that can replace each other in some class of contexts with insignificant changes of the whole text's meaning." (From "Synonymy".)

See also synonym

abstract models

Abstract models describe structures commonly found in resource descriptions and other information resources, regardless of the specific domain. (From "Structuring Descriptions".)

accessioning

Adding a resource to a library collection is called *acquisition*, but adding to a museum collection is called *accessioning*. (From "Introduction".)

See also acquisition, collection development

accuracy

See precision.

ACM

Association for Computing Machinery (ACM)

(`http://www.acm.org/`)

acquisition

Adding a resource to a library collection is called *acquisition*, but adding to a museum collection is called *accessioning*. (From "Introduction".)

See also accessioning, collection development

active resources

Active resources create effects or value on their own, sometimes when they initiate interactions with passive resources. Active resources can be people, other living resources, computational agents, active information sources, web-based services, self-driving cars, robots, appliances, machines or otherwise ordinary objects like light bulbs, umbrellas, and shoes that have been made "smarter." (From "Active or Operant Resources".)

activities

There are four *activities* that occur naturally in every *organizing system*; how explicit they are depend on the scope, the breadth or variety of the resources, and the scale, the number of resources that the organizing system encompasses. (From "Introduction".)

See also selecting, organizing, designing resource-based interactions, maintaining

ad hoc category

An **ad hoc category** or goal-derived category is a collection of resources that happen to go together to satisfy a goal. The resources might not have any discernible properties in common. (From "Principles for Creating Categories".)

affordance

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. (From "Affordance and Capability".)

See also capability

agency

Agency is the extent to which a resource can initiate actions on its own. We can define a continuum between completely passive resources that cannot initiate any actions and active resources that can initiate actions based on information they sense from their environments or obtain through interactions with other resources. (From "Resource Agency".)

agent

We use the more general word, *agent*, for any entity capable of autonomous and intentional organizing effort, because it treats organizing work done by people and organizing work done by computers as having common goals, despite obvious differences in methods. (From "The Concept of "Agent"".)

agents

A facet in the hierarchical structure of the AAT thesaurus. Basically, people and the various groups and organizations with which they identify, whether based on physical, mental, socio-economic, or political characteristics—e.g., "stonemasons" or "socialists." (From "Faceted Classification in Description".)

alias

See synonym.

alphabetical ordering

Alphabetical ordering is arranging resources according to their names (From "The Concept of "Organizing Principle"".)

See also chronological ordering

American Society for Information Science and Technology

See ASIS&T.

analysis

A common interaction with an organizing system.

analytico-synthetic classification

In library science a classification system that builds categories by combination of facets is sometimes also called *analytico-synthetic*. (From "Classification Schemes".)

anchor text

In web contexts, the words in which a structural link is embedded are called the *anchor text*. (From "Hypertext Links".)

See also hypertext

ANSI

American National Standards Association (ANSI)

(http://www.ansi.org/)

antonymy

Antonymy is the lexical relationship between two words that have opposite meanings. *Antonymy* is a very salient lexical relationship, and for adjectives it is even more powerful than synonymy. (From "Antonymy".)

APA

American Psychiatric Association (APA)

(http://www.psych.org)

API

application program interfaces (APIs)

appraisal

What is the value of this resource? What is its cost? At what rate does it depreciate? Does it have a shelf life? Does it have any associated ratings, rankings, or quality measures? Moreover, what is the quality of those ratings, rankings and measures? (From "Resource Description to Support Selection".)

architectural perspective

The architectural perspective emphasizes the number and abstraction level of the components of a relationship, which together characterize its complexity. (From "Describing Relationships: An Overview".)

arity

The *degree* or *arity* of a relationship is the number of entity types or categories of resources in the relationship. This is usually, though not always, the same as the number of arguments in the relationship expression. (From "Degree".)

See also degree

ARPA

Advanced Research Projects Agency

(`http://www.darpa.mil/`)

artifact

See resource.

ASCII

American Standard Code for Information Interchange (ASCII)

American National Standard for Information Systems—Coded Character Sets—7-Bit American National Standard Code for Information Interchange (7-Bit ASCII), ANSI X3.4-1986, American National Standards Institute, Inc., March 26, 1986

asset

See resource.

ASIS&T

American Society for Information Science and Technology (ASIS&T)

(`http://www.asis.org`)

associated resource

See description resources.

associative array

See dictionary.

asymmetric relationships

Asymmetric relationships express a one-way relationship from the subject to the object. For example, "is-parent-of."

See also hypertext, directionality, and one-way link. From asymmetric relationship.)

attribute

Attribute is a synonym for "*property*."

To **attribute** is to assert or assign a value to a property. See attribution relationship

An **attribute** is a syntactic component of XML elements and a conceptual component of the XML Infoset, consisting of a potentially qualified name and a value, whose type may influence its interpretation. The value of an attribute in an XML document is a Unicode string. The value of that attribute in the XML Infoset could be a simple string of text, a precisely-typed numeric or temporal value, a list of references to document nodes, a hypertext link, or a reference to a formal notation. (See also element item)

attribution relationship

Asserting or assigning values to properties; the predicate depends on the property: "is-the-author-of," "is-married-to," "is-employed-by," etc. (From "Types of Semantic Relationships".)

authentication

Is the resource what it claims to be? ("Authenticity") Resource descriptions that can support authentication include technological ones like time stamps, watermarking, encryption, checksums, and digital signatures. (From "Determining the Purposes".)

authenticity

In ordinary use we say that something is *authentic* if it can be shown to be, or has come to be accepted as what it claims to be. The importance and nuance of questions about authenticity can be seen in the many words we have to describe the relationship between "the real thing" (the "original") and something else: copy, reproduction, replica, fake, phony, forgery, counterfeit, pretender, imposter, ringer, and so on. (From "Authenticity".)

See also provenance

authority control

For bibliographic resources important aspects of vocabulary control include determining the authoritative forms for author names, uniform titles of works, and the set of terms by which a particular subject will be known. In library science, the process of creating and maintaining these standard names and terms is known as *authority control*. (From "Use Controlled Vocabularies".)

B

BI

Business Intelligence (BI)

bi-directional

See symmetric relationships.

bi-directional links

When a *bi-directional link* is created between an anchor and a destination, it is as though a one-way link that can be followed in the opposite direction is automatically created. Two one-way links serve the same purpose, but the return link is not automatically established when the first one is created. (From "Hypertext Links".)

See also hypertext, directionality, one-way link

bibliographic description

Bibliographic descriptions characterize information resources and the entities that populate the bibliographic universe, which include works, editions, authors, and subjects. (From "Bibliographic Descriptions".)

bibliography

A **bibliography** is a description resource in the domain of library science. (Ed.)

bibliometrics

Information scientists began studying the structure of scientific citation, now called *bibliometrics*, nearly a century ago to identify influential scientists and publications. (From "Bibliometrics, Shepardizing, Altmetrics, and Social Network Analysis".)

big data

For digital resources, inexpensive storage and high bandwidth have largely eliminated capacity as a constraint for organizing systems, with an exception for *big data*, which is defined as a collection of data that is too big to be managed by typical database software and hardware architectures. (From "Architectural Thinking".)

binary antonyms

Contrasting or *binary antonyms* are used in mutually exclusive contexts where one or the other word can be used, but never both. For example, "alive" and "dead" can never be used at the same time to describe the state of some entity, because the meaning of one excludes or contradicts the meaning of the other. (From "Antonymy".)

binary link

A **binary link** connects one anchor to one destination. (From the Sidebar, "Perspectives on Hypertext Links".)

See also hypertext

BISAC

Book Industry Standards Advisory Committee classification (BISAC)

BISG

Book Industry Study Group (BISG)

(`http://www.bisg.org/`)

blob

A **blob** is any resource whose internal structure is functionally opaque for the purpose at hand. (From "Blobs".)

Boolean facet

Take on one of two values, yes (true) or no (false) along some dimension or property. (From Boolean facets.)

See also "A Classification for Facets"

born digital

Resources in organizing systems that are created in digital format are **born digital**. These include resources created by word processors and digital cameras, or by audio and video recorders. Other resources are produced in digital form by "smart things" and by the systems that create digital resources when they interact with barcodes, QR ("quick response") codes, RFID tags, or other mechanisms for tracking identity and location. (From "Resource Format".)

BPEL

Business Process Execution Language (BPEL)

(https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=wsbpel)

C

CAFE

Corporate Average Fuel Economy (CAFE)

(http://www.nhtsa.gov/staticfiles/rulemaking/pdf/cafe/CAFE-GHG_MY_2012-2016_Final_Rule_FR.pdf)

capability

Capability is a function of the affordances offered by an *organizing system* and the possible interactions they imply. (From capability.)

See also affordance

capability and compatibility

Will the resource meet functional or interoperability requirements? Technology-intensive resources often have numerous specialized types of descriptions that specify their functions, performance, reliability, and other "-ilities" that determine if they fit in with other resources in an organizing system. (From "Resource Description to Support Selection".)

cardinality

The *cardinality* of a relationship is the number of instances that can be associated with each entity type in a relationship. (From "Cardinality".)

cataloging

Documenting the contents of library and museum collections to organize them is called *cataloging* (From "Introduction".)

See also collection development

categories

Categories are *equivalence classes*, sets or groups of things or abstract entities that we treat the same. (From "The What and Why of Categories".)

See also equivalence class

CBS

CBS Corporation and CBS Broadcasting Inc.

(`http://www.cbs.com/`)

CC

Common Cartridge and Learning Tools Interoperability

(`http://www.imsglobal.org/commoncartridge.html`)

CDWA

Categories for the Description of Works of Art (CDWA)

(`http://www.getty.edu/research/publications/electronic_publications/cdwa/`)

centrality

The **centrality** of a resource instance as a member of a category is a measure of how close it is to a mathematical average on some measures or property values that apply to all the members. (From "Probabilistic Categories and "Family Resemblance"".)

CERN

European Organization for Nuclear Research (*Centre Européen de Research Nuclear*)

(http://public.web.cern.ch/public/)

character

Unicode makes the important distinction between *characters* and *glyphs*. A *character* is the smallest meaningful unit of a written language. In alphabet-based languages like English, *characters* are letters; in languages like Chinese, characters are ideographs. (From “Notations”.)

character encoding

A notation that has had numbers assigned to its characters is called a *character encoding*. (From “Notations”.)

The most ambitious character coding in existence is Unicode, which as of version 6.0 assigns numbers to 109,449 characters. Unicode makes the important distinction between characters and glyphs.

chronological ordering

Chronological ordering is arranging resources according to the date of their creation or other important event in the lifetime of the resource (From “The Concept of “Organizing Principle””.).

See also alphabetical ordering

CIDR

Classless Inter-Domain Routing (CIDR)

circulation

We might treat *circulation*, borrowing and returning the same item, as one of the interactions with resources that defines a library. (From “The Concept of “Collection””.)

See also collection development

classes

In object-oriented programming languages, *classes* are schemas that serve as templates for the creation of objects. A class in a programming language is analogous to a database schema that specifies the structure of its member instances, in that the class definition specifies how instances of the class are constructed in terms of data types and possible values. Programming classes may also specify whether data in a member object can be accessed, and if so, how. (From “Implementing Categories Defined by Properties”.)

classical categories

Categories defined by necessary and sufficient properties are also called *monothetic*. They are also sometimes called *classical categories* because they conform to Aristotle’s theory of how categories are used in logical deduction using syllogisms. (From “Necessary and Sufficient Properties”.)

classification

The systematic assignment of resources to a system of intentional categories, often institutional ones. (From “Classification vs. Categorization”.)

Classification is applied categorization – the assignment of resources to a system of categories, called classes, using a predetermined set of principles.

See also inclusion

classification scheme

See classifications

classifications

A system of categories and its attendant rules or access methods is typically called a *classification scheme* or just the *classifications*. A system of categories captures the distinctions and relationships among its resources that are most important in a domain and for a particular context of use, creating a reference model or conceptual roadmap for its users. (From "Introduction".)

classifying

When we make an assertion that a particular instance is a member of class, we are *classifying* the instance. (From "Inclusion".)

class inclusion

Class inclusion is the fundamental and familiar "**is-a,**" "**is-a-type-of**," or "**subset**" relationship between two entity types or classes where one is contained in and thus more specific than the other more generic one. (From "Inclusion".)

See also inclusion

clustering

Clustering techniques share the goal of creating meaningful categories from a collection of items whose properties are hard to directly perceive and evaluate, which implies that category membership cannot easily be reduced to specific property tests and instead must be based on similarity. (From "Categories Created by Clustering".)

The end result of **clustering** is a statistically optimal set of categories in which the similarity of all the items within a category is larger than the similarity of items that belong to different categories.

cognates

Many words in different languages have common roots, and as a result are often spelled the same or nearly the same. This is especially true for technology words; for example, "computer" has been borrowed by many languages. The existence of these *cognates* and borrowed words makes us vulnerable to false cognates. When a word in one language has a different meaning and refers to different resources in another, the results can be embarrassing or disastrous. "*Gift*" is poison in German; "*pain*" is bread in French. (From "Homonymy, Polysemy, and False Cognates".)

collection

A *collection* is a group of resources that have been selected for some purpose. (From "The Concept of "Collection"".)

collection development

Libraries and museums usually make their *selection* principles explicit in *collection development* policies. 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*. *Circulation* is a central interaction in libraries, but because museum resources do not circulate the primary interactions for museum users are *viewing* or *visiting* the collection. *Maintenance* activities are usually described as *preservation* or *curation*. (From "Introduction".)

collocation

The Organizing System for a small collection can sometimes use only the minimal or default organizing principle of *colocation*—putting all the resources in the same location: in the same container, on the same shelf, or in the same email in-box. (From “The Concept of “Organizing Principle””.)

compliance

Compliance is a maintenance activity.

component-object inclusion

Component-Object is the relationship type when the part is a separate component that is arranged or assembled with other components to create a larger resource. (From “Types of Semantic Relationships”.)

See also inclusion

compounding

Putting two “free morphemes” together. (From “Derivational Morphology”.)

constraint

A limit or bound on a data type or structure, most usefully expressed in a schema or regular expression. Constraints on data types and structures can be expressed in a variety of natural, programming and schema languages with varying degrees of efficacy. (Ed.)

content rules

Content rules are similar to controlled vocabularies because they also limit the possible values that can be used in descriptions. Instead of specifying a fixed set of values, content rules typically restrict descriptions by requiring them to be of a particular data type (integer, Boolean, Date, and so on). (From “Controlled Vocabularies and Content Rules”.)

contextual properties

Contextual properties are those related to the situation or context in which a resource is described. Dey defines *context* as "any information that characterizes a situation related to the interactions between users, applications, and the surrounding environment." (From "Extrinsic Dynamic Properties".)

controlled vocabulary

One way to encourage good names for a given resource domain or task is to establish a *controlled vocabulary*. A *controlled vocabulary* is like a fixed or closed dictionary that includes the terms that can be used in a particular domain. A controlled vocabulary shrinks the number of words used, reducing synonymy and homonymy, eliminating undesirable associations, leaving behind a set of words with precisely defined meanings and rules governing their use. (From "Use Controlled Vocabularies".)

coverage

The values of a facet should be able of classifying all instances within the intended scope. (From "Design Principles and Pragmatics".)

CRM

Customer Relationship Management (CRM)

crosswalk

Similar to mapping, a straightforward approach to transformation is the use of *crosswalks*, which are equivalence tables that relate resource description elements, semantics, and writing systems from one organizing system to those of another. (From "Modes of Transformation".)

cultural categories

Cultural categories are the archetypical form of categories upon which individual and institutional categories are usually based. Cultural categories tend to describe our everyday experiences of the world and our accumulated cultural knowledge. (From “Cultural Categories”.)

cultural properties

Cultural properties derive from conventional language or culture, often by analogy, because they can be highly evocative and memorable. (From “Extrinsic Dynamic Properties”

curation

Curation is a maintenance activity.

Curation usually refers to the methods or systems that add value to and preserve resources, while the concept of *governance* more often emphasizes the institutions or organizations that carry out those activities. The former is most often used for libraries, museums, or archives and the latter for enterprise or inter-enterprise contexts. (From “When Is It Being Organized?”.)

See also collection development

D

data

Data is a collection of one or more pieces if information. The singular noun form is "datum"; the plural forms are "datums" and "data"; the collective noun form is also "data". For example: Starting with a single datum; many more datums are subsequently identified; those data are then intentionally arranged; and, finally the data is organized.

data activities

Data capture, extraction and generation are resource selection activities.

Data cleaning and cleansing are maintenance activities.

Data insertion and integration add resources to a collection.

data rot

Data rot is a colloquial term intended to convey the fact that the physical medium of a digital resource deteriorates over time.

data schema

Data schemas that specify data entities, elements, identifiers, attributes, and relationships in databases and XML document types on the transactional end of the Document Type Spectrum ("Resource Domain") are implementations of the categories needed for the design, development and maintenance of information organization systems. Data schemas tend to rigidly define categories of resources. (From "Implementing Categories Defined by Properties".)

data science

Data science, actuarial science, statistics, probability, and predictive analytics. Predicting future outcomes by applying statistical analysis over many large datasets and calculating probabilities. Ancient roots in the fields of economics, insurance, cartography, astronomy, and meteorology.

DC

Dublin Core (DC)

(http://dublincore.org/documents/dcmi-terms/)

See also [Hillmann2005]

DCMI

Dublin Core Metadata Initiative (DCMI)

(http://dublincore.org/)

DDC

Dewey Decimal Classification (DDC)

(http://www.oclc.org/dewey/)

decision tree

A simple *decision tree* is an algorithm for determining a decision by making a sequence of logical or property tests. (From "Implementing Categories Defined by Properties".)

decoding

A digital resource is first a sequence of bits. Decoding transforms those bits into characters according to the encoding scheme used, extracting the text from its stored form.

degree

The *degree* or *arity* of a relationship is the number of entity types or categories of resources in the relationship. This is usually, though not always, the same as the number of arguments in the relationship expression. (From “Degree”.)

derivational morphology

Derivational morphology deals with how words are created by combining morphemes. (From “Derivational Morphology”.)

description resources

Any primary resource can have one or more *description resources* associated with it to facilitate finding, interacting with, or interpreting the primary one. *Description resources* are essential in organizing systems where the primary resources are not under its control and can only be accessed or interacted with through the description. *Description resources* are often called *metadata*. (From “Resource Focus”.)

Description resources, such as physical or online catalog records, describe the *primary resources* that comprise the collection.

descriptive control

Descriptive control is objective and straightforward, *lining up a population of writings in any arbitrary order.* (From “Bibliographic Descriptions”.)

descriptor

In the library science context of *bibliographic description,* a *descriptor* is one of the terms in a carefully designed language that can be assigned to a resource to designate its properties, characteristics, or meaning, or its relationships with other resources. (From ““Description” as an Inclusive Term”.)

designed resource access policies

Designed resource access policies are established by the designer or operator of an organizing system to satisfy internally generated requirements. (From “Introduction”.)

designing resource-based interactions

Designing and implementing the actions, functions or services that make use of the resources. (From “Introduction”.)

dictionary

A *dictionary* is a set of property-value pairs or entries. It is a set of entries, not a list of entries, because the pairs are not ordered and because each entry must have a unique key.

Note that this specialized meaning of *dictionary* is different from the more common meaning of “dictionary” as an alphabetized list of terms accompanied by sentences that define them. (From “Dictionaries”.)

digitization

Other digital resources are created by *digitization*, the process for transforming an artifact whose original format is physical so it can be stored and manipulated by a computer. (From “Resource Format”.)

dimensionality reduction

Dimensionality reduction implies transforming a high-dimensional space into a lower-dimensional one. Reducing the number of components in a multidimensional description can be accomplished by many different statistical techniques that go by names like “feature extraction,” “principle components analysis,” “orthogonal decomposition,” “latent semantic analysis,” “multidimensional scaling,” and “factor analysis.” (From “Vocabulary Control as Dimensionality Reduction”.)

directionality

The *directionality* of a relationship defines the order in which the arguments of the relationship are connected. A *one-way* or *uni-directional* relationship can be followed in only one direction, whereas a *bi-directional* one can be followed in both directions. (From “Directionality”.)

See also hypertext, directionality, one-way link, bi-directional

discipline

A *discipline* is an integrated field of study in which there is some level of agreement about the issues and problems that deserve study, how they are interrelated, how they should be studied, and how findings or theories about the issues and problems should be evaluated. (From “The Discipline of Organizing”.)

discovery

What available resources might be added to a collection? New resources are often listed in directories, registries, or catalogs. Some types of resources are selected and acquired automatically through subscriptions or contracts. (From “Determining the Purposes”.)

DNS

Domain Name System (DNS)

(`http://tools.ietf.org/html/rfc1035`)

document

See resource.

document frequency

Inverse document frequency (idf) is a collection-level property. The *document frequency* (df) is the number of resources containing a particular term. The inverse document frequency (idf) for a term is defined as $idf_t = \log(N/df_t)$, where N is the total number of documents. The inverse document frequency of a term decreases the more documents contain the term, providing a discriminating factor for the importance of terms in a query. (From "Ranked Retrieval with Vector Space or Probabilistic Models".)

DOI

Digital Object Identifier (DOI)

(http://www.doi.org)

domain

Resource domain is an intuitive notion that groups resources according to the set of natural or intuitive characteristics that distinguishes them from other resources. It contrasts with the idea of ad hoc or arbitrary groupings of resources that happen to be in the same place at some time. (From "Resource Domain".)

DPLA

Digital Public Library of America (DPLA)

(http://dp.la/)

DRM

digital rights management (DRM)

DSM

Diagnostic and Statistical Manual of Mental Disorders (DSM)

(http://www.dsm5.org/)

DTD

Document Type Definition (DTD)

E

ECM

Enterprise Content Management (ECM)

edge

See tree.

EDI

Electronic Data Exchange (EDI)

Typically refers to one or all of the UN/EDIFACT, ANSI ASC X12, TRADCOMS and ODETTE standards.

EDM

Enterprise Data Management (EDM)

effectivity

Many resources, or their properties, also have locative or temporal *effectivity*, meaning that they come into effect at a particular time and/or place; will almost certainly cease to be effective at some future date, and may cease to be effective in different places. (From "Effectivity".)

element item

An *element* item has a set of *attribute* items, and a list of child nodes. These child nodes may include other element items, or they may be character items. (From "XML Information Set".)

encoding scheme

An *encoding scheme* is a specialized *writing system* or syntax for particular types of values. (From "Controlling Values".)

energy facet

One of Ranganathan's universal facets in colon classification. The action or activity of the thing. (From "Foundations for Faceted Classification".)

entity

See resource.

entity type

See classes

enumeration

The simplest principle for creating a category is *enumeration*; any resource in a finite or countable set can be deemed a category member by that fact alone. (From "Enumeration".)

See also extensional definition.

enumerative classification

Classification schemes in which all possible categories to which resources can be assigned are defined explicitly are *enumerative*. (From "Classification Schemes".)

enumerative facets

Have mutually exclusive possible values. (From “A Classification for Facets”.)

equivalence class

See categories

equivalence relationship

Any relationship that is both symmetric and transitive is an *equivalence relationship*; “is-equal-to” is obviously an equivalence relationship because if A=B then B=A and if A=B and B=C, then A=C. Other relationships can be equivalent without meaning “exactly equal,” as is the relationship of “is-congruent-to” for all triangles. (From “Equivalence”.)

ERP

Enterprise Resource Planning (ERP)

ETL

Extract, Transform, and Load

EXIF

Exchangeable Image File Format (EXIF)

(`http://www.exif.org/`)

expression

The distinctions put forth by Panizzi, Lubetzky, Svenonius and other library science theorists have evolved today into a four-step abstraction hierarchy (see Figure: The FRBR Abstraction Hierarchy.) between the abstract *work*, an *expression* in multiple formats or genres, a particular *manifestation* in one of those formats or genres, and a specific physical *item*. (From "Identity and Bibliographic Resources".)

exploitive control

Exploitative control, defined as *the ability to make the best use of a body of writings*, requires descriptions that evaluate resources for their suitability for particular uses. (From "Bibliographic Descriptions".)

extensibility of classification

See flexibility

extension

See extensional definition

extensional definition

The simplest principle for creating a category is *enumeration*; any resource in a finite or countable set can be deemed a category member by that fact alone. This principle is also known as *extensional definition*, and the members of the set are called the *extension*. (From "Enumeration".)

F

faceted classification

In a *faceted classification* system, each resource is described using properties from multiple facets, but a person searching for resources does not need to consider all of the properties (and consequently the facets) and does not need to consider them in a fixed order, which an enumerative hierarchical classification requires. (From “Faceted Classification”.)

family resemblance

A second consequence is that the sharing of some but not all properties creates what we call *family resemblances* among the category members; just as biological family members do not necessarily all share a single set of physical features but still are recognizable as members of the same family. (From “Probabilistic Categories and “Family Resemblance””.)

FCC

Federal Communications Commission (FCC)

(`http://www.fcc.gov/`)

FDA

Food and Drug Administration (FDA)

(`http://www.fda.gov/`)

feature

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. (From “Organizing Resources”.)

See also property

feature-activity inclusion

Feature-Activity is a relationship type in which the components are stages, phases, or sub activities that take place over time.

This relationship is similar to component-object in that the components in the whole are arranged according to a structure or pattern. (From "Types of Semantic Relationships".)

See also inclusion

FERPA

Family Educational Rights and Privacy Act (FERPA)

(`http://www2.ed.gov/policy/gen/guid/fpco/ferpa/`)

finding

What resources are available that "correspond to the user's stated search criteria" and thus can satisfy an information need? Before there were online catalogs and digital libraries, we found resources by referencing catalogs of printed resource descriptions incorporating the title, author, and subject terms as access points into the collection; the subject descriptions were the most important finding aids when the user had no particular resource in mind. Modern users accept that computerized indexing makes search possible over not only the entire description resource, but often over the entire content of the primary resource. (From "Resource Description to Support Interactions".)

flexibility of classification

A related principle about maintaining classifications over time is *flexibility*, the degree to which the classification can accommodate new categories. Computer scientists typically describe this principle as *extensibility*, and library scientists sometimes describe it as *hospitality*. (From "Principles for Maintaining the Classification over Time".)

FOAF

Friend of a Friend (FOAF)

(http://www.foaf-project.org/)

focus

The contrast between *primary resources* and *description resources* is very useful in many contexts, but when we look more broadly at organizing systems, it is often difficult to distinguish them, and determining which resources are primary and which are *metadata* is often just a decision about which resource is currently the *focus* of our attention. (From “Resource Focus”.)

fonds

The “original order” of the resources in an archive embodies the implicit or explicit organizing system of the person or entity that created the documents; 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, preserving any hierarchy of boxes, folders, envelopes, and individual documents—and thus they are not re-organized according to other (perhaps more systematic) classifications. (From “What Is Being Organized?”.)

font

A *font* is a collection of glyphs used to depict some set of *characters*. A Unicode font explicitly associates each glyph with a particular number in the Unicode character encoding. (From “Notations”.)

form

We treat the set of implementation decisions about character notations, syntax, and structure as the *form* of resource description (From “Frameworks for Resource Description”.)

format

Information resources can exist in numerous **formats** with the most basic format distinction being whether the resource is physical or digital.

FRAD

Functional Requirements for Authority Data (FRAD)

(http://www.ifla.org/publications/functional-requirements-for-authority-data)

framework

A *framework* is a set of concepts that provide the basic structure for understanding a domain, enabling a common vocabulary for different explanatory theories. (From "The Discipline of Organizing".)

FRBR

Functional Requirements for Bibliographical Records (FRBR)

(http://www.ifla.org/publications/functional-requirements-for-bibliographic-records)

frequency of use principle

Some organization emerges implicitly through a *frequency of use* principle. In your kitchen or clothes closet, the resources you use most often migrate to the front because that is the easiest place to return them after using them. (From "The Concept of "Organizing Principle"".)

FRSAD

Functional Requirements for Subject Authority Data (FRSAD)

(http://www.ifla.org/files/assets/classification-and-indexing/functional-requirements-for-subject-authority-data/frsad-final-report.pdf)

FTC

Federal Trade Commission (FTC)

(http://www.ftc.gov/)

FTP

File Transfer Protocol (FTP)

(http://tools.ietf.org/html/rfc959)

G

globally unique identifier (GUID)

A **globally unique identifier** (or GUID), is an identifier that will never be the same as another identifier in any organizing system anywhere else. (Ed.)

glossary definition

A glossary definition states the meaning of its corresponding term. There must be one and there may be more definitions for a given term. The most common presentation is a set of words or symbols that convey the semantic of the term, such as the expanded form of an abbreviation or acronym, or a paragraph of text. Definition by reference is often used for synonym terms.

See also synonym

glyph

A specific mark that can be used to depict a *character* is a *glyph*. (From "Notations".)

See also character, font

governance

Curation usually refers to the methods or systems that add value to and preserve resources, while the concept of *governance* more often emphasizes the institutions or organizations that carry out those activities. The former is most often used for libraries, museums, or archives and the latter for enterprise or inter-enterprise contexts. (From "When Is It Being Organized?".)

GPS

Global Positioning System

(http://www.schriever.af.mil/GPS/)

gradience

When category members differ in their centrality or typicality to the category definition, this effect is called category gradience. (From "Probabilistic Categories and "Family Resemblance"".)

grammar

The *syntax* and *grammar* of a language consists of the rules that determine which combinations of its words are allowed and are thus grammatical or *well-formed*. Natural languages have substantial similarities by having nouns, verbs, adjectives and other parts of speech, but they differ greatly in how they arrange them to create sentences. (From "Syntax and Grammar".)

granularity

Granularity refers to the level of detail or precision for a specific information resource property. For instance, the postal address of a particular location might be represented as several different data items, including the number, street name, city, state, country and postal code (a high-granularity model). It might also be represented in one single line including all of the information above (a low-granularity model). (From "Granularity and Abstraction".)

graph

Like a *tree*, a *graph* consists of a set of nodes connected by edges. These edges may or may not have a direction ("Directionality"). If they do, the *graph* is referred to as a "directed graph." If a *graph* is directed, it may be possible to start at a node and follow edges in a path that leads back to the starting node. Such a path is called a "cycle." If a directed graph has no cycles, it is referred to as an "acyclic graph." (From "Graphs".)

GUID

Globally Unique Identifier

H

hierarchical classification

When multiple resource properties are considered in a fixed sequence, each property creates another level in the system of categories and the classification scheme is *hierarchical* or *taxonomic*. (From "Classification Schemes".)

hierarchical facet

Organize resources by logical inclusion ("Inclusion"). (From "A Classification for Facets".)

See also inclusion

HIPAA

Health Insurance Portability and Accountability Act (HIPAA)

(`http://www.hhs.gov/ocr/privacy/hipaa/understanding/`)

homographs

When two words are spelled the same but have different meanings they are *homographs*; if they are also pronounced the same they are *homonyms*. If the different meanings of the homographs are related, they are *polysemes*. (From "Homonymy, Polysemy, and False Cognates".)

homonyms

Homonyms are homographs that are pronounced the same. (From "Homonymy, Polysemy, and False Cognates".)

hospitality of classification

See flexibility

HR

Human Resources

HTML

Hypertext Markup Language (HTML)

(`http://www.w3.org/community/webed/wiki/HTML/Specifications`)

HTTP

Hypertext Transfer Protocol (HTTP)

(`http://www.w3.org/Protocols/`)

hypernym

When words encode the semantic distinctions expressed by class inclusion, the word for the more specific class in this relationship is called the *hyponym*, while the word for the more general class to which it belongs is called the *hypernym*. (From "Hyponymy and Hyperonymy".)

hypertext

Hypertext expresses relationships among resources. Hypertext is "a provision whereby any item may be caused at will to select immediately and automatically another." It can be used to create non-sequential narratives that gives choices to readers. (See "Hypertext Links".)

hypertext link

The concept of read-only or follow-only structures that connect one document to another is usually attributed to Vannevar Bush in his seminal 1945 essay titled *As We May Think.* Bush called it *associative indexing,* defined as “a provision whereby any item may be caused at will to select immediately and automatically another.” (From “Hypertext Links”.)

hyponym

When words encode the semantic distinctions expressed by class inclusion, the word for the more specific class in this relationship is called the *hyponym*, while the word for the more general class to which it belongs is called the *hypernym*. (From “Hyponymy and Hyperonymy”.)

I

IAU

International Astronomical Union (IAU)

(`http://www.iau.org/`)

IBM

International Business Machines (IBM)

(`http://www.ibm.com`)

ICANN

Internet Corporation for Assigned Names and Numbers (ICANN)

(http://www.icann.org/)

ICD-10-CM

International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM)

(http://www.cdc.gov/nchs/icd/icd10cm.htm)

identifier

An *identifier* is a special kind of name assigned in a controlled manner and governed by rules that define possible values and naming conventions. (From "Identity, Identifiers, and Names".)

identifying

Another purpose of resource description is to enable a user to confirm the identity of a specific resource or to distinguish among several that have some overlapping descriptions. In bibliographic contexts this might mean finding the resource that is identified by its citation. Computer processable resource descriptions like bar codes, QR codes, or RFID tags are also used to identify resources. In Semantic Web contexts, URIs serve this purpose. (From "Resource Description to Support Interactions".)

identity

When some thing or things are treated as a single resource this establishes an **identity**. (Ed.)

IEEE

Institute of Electrical and Electronics Engineers

(http://www.ieee.org/index.html)

IETF

Internet Engineering Task Force

(http://ietf.org)

IFLA

International Federation of Library Associations and Institutions (IFLA)

(http://www.ifla.org/)

IHTSDO

International Health Terminology Standards Development Organization (IHTSDO)

(http://www.ihtsdo.org/)

implementation perspective

The implementation perspective considers how the relationship is implemented in a particular notation and syntax and the manner in which relationships are arranged and stored in some technology environment. (From "Describing Relationships: An Overview".)

implicit classification

Because names and dates can take on a great many values, an organizing principle like *alphabetical* or *chronological* ordering is unlikely to enumerate in advance an explicit category for each possible value. Instead, we can consider these organizing principles as creating an *implicit or latent* classification system in which the categories are generated only as needed. For example, the Q category only exists in an alphabetical scheme if there is a resource whose name starts with Q. (From "Classification Schemes".)

imposed policies

Imposed Policies are mandated by an external entity and the organizing system must comply with them. (From "Access Policies".)

inclusion relationship

One entity type contains or is comprised of other entity types; often expressed using "is-a," "is-a-type-of," "is-part-of," or "is-in" predicates. (From "Types of Semantic Relationships".)

See also component-object, feature-activity inclusion, locative, member-collection, meronymic, part-whole, phase-activity, place-area, portion-mass, stuff-object, temporal, topological, taxonomy and classification

index

An *index* is a *description resource* that contains information about the locations and frequencies of terms in a document *collection* to enable it to be searched efficiently. (From "The Concept of "Collection"".)

individual categorization

Individual categories are created in an organizing system to satisfy the *ad hoc* requirements that arise from a person's unique experiences, preferences, and resource collections. Unlike cultural categories, which usually develop slowly and last a long time, individual categories are created by intentional activity, in response to a specific situation, or to solve an emerging organizational challenge. (From "Individual Categories".)

inflectional morphology

Inflectional mechanisms change the form of a word to represent tense, aspect, agreement, or other grammatical information. Unlike derivation, inflection never changes the part-of-speech of the base morpheme. The *inflectional morphology* of English is relatively simple compared with other languages. (From “Inflectional Morphology”.)

informatics

Informatics is a broad academic category encompassing the science of information, including the automation of information processing. Computer science, information architecture and web architecture are among the related academic disciplines.

information architecture

Abstract patterns of information content or organization are sometimes called architectures, so it is straightforward from the perspective of the discipline of organizing to define the activity of *information architecture* as designing an abstract and effective organization of information and then exposing that organization to facilitate navigation and information use. (From ““Information Architecture” and Organizing Systems”.)

information component

An *information component* can be: (1) Any piece of information that has a unique label or identifier or (2) Any piece of information that is self-contained and comprehensible on its own. (From “Identity and Information Components”.)

information organization

Traditional information organization activities include bibliographic description and cataloging.

information retrieval

Traditional information retrieval activities include automated text processing, indexing and search.

instance

See resource.

institutional categorization

In contrast to cultural categories that are created and used implicitly, and to individual categories that are used by people acting alone, *institutional categories* are created and used explicitly, and most often by many people in coordination with each other. Institutional categories are most often created in abstract and information-intensive domains where unambiguous and precise categories are needed to regulate and systematize activity, to enable information sharing and reuse, and to reduce transaction costs. (From "Institutional Categories".)

institutional semantics

Systems of *institutional semantics* offer precisely defined abstractions or *information components* ("Identity and Information Components") needed to ensure that information can be efficiently exchanged and used. (From "Institutional Semantics".)

institutional taxonomies

Institutional taxonomies are classifications designed to make it more likely that people or computational agents will organize and interact with resources in the same way. (From "Institutional Taxonomies".)

integration

Integration is the controlled sharing of information between two (or more) business systems, applications, or services within or between firms. *Integration* means that one party can extract or obtain information from another one, it does not imply that the recipient can make use of the information. (From "Integration and Interoperability".)

integrity of classification

Changes in the meaning of the categories in a classification threaten its *integrity*, the principle that categories should not move within the structure of the classification system. (From "Principles for Maintaining the Classification over Time".)

intension

Categories whose members are determined by one or more properties or rules follow the principle of *intensional definition*, and the defining properties are called the *intension*. (From "Single Properties".)

intensional definition

Categories whose members are determined by one or more properties or rules follow the principle of *intensional definition*, and the defining properties are called the *intension*. (From "Single Properties".)

intentional arrangement

Intentional arrangement emphasizes explicit or implicit acts of organization by people, or by computational processes acting as proxies for, or as implementations of, human intentionality. (From "The Concept of "Intentional Arrangement"".)

interaction

An *interaction* is an action, function, service, or capability that makes use of the resources in a collection or the collection as a whole. The interaction of *access* is fundamental in any collection of resources, but many Organizing Systems provide additional functions to make access more efficient and to support additional interactions with the accessed resources. (From "The Concept of "Interactions"".)

interoperability

Interoperability goes beyond integration to mean that systems, applications, or services that exchange information can make sense of what they receive. *Interoperability* can involve identifying corresponding components and relationships in each system, transforming them syntactically to the same format, structurally to the same granularity, and semantically to the same meaning. (From "Integration and Interoperability".)

intrinsic meaning interpretation

One of Panofsky's three levels of description for artistic resources. At this level, context and deeper understanding come into play—including what the creator of the description knows about the situation in which the work was created. Why, for example, did this particular artist create this particular depiction of The Last Supper in this way? Panofsky posited that professional art historians are needed here, because they are the ones with the education and background necessary to draw meaning from a work. (From "Describing Museum and Artistic Resources".)

inverse document frequency

Inverse document frequency (idf) is a collection-level property. The *document frequency* (df) is the number of resources containing a particular term. The inverse document frequency (idf) for a term is defined as $idf_t = \log(N/df_t)$, where N is the total number of documents. The inverse document frequency of a term decreases the more documents contain the term, providing a discriminating factor for the importance of terms in a query. (From “Ranked Retrieval with Vector Space or Probabilistic Models”.)

inverse relationship

For asymmetric relationships, it is often useful to be explicit about the meaning of the relationship when the order of the arguments in the relationship is reversed. The resulting relationship is called the *inverse* or the converse of the first relationship. (From “Inverse”.)

ISBN

International Standard Book Number (ISBN)

(http://www.isbn.org/)

ISO

International Organization for Standardization (ISO)

(http://www.iso.org/iso/)

item

The distinctions put forth by Panizzi, Lubetzky, Svenonius and other library science theorists have evolved today into a four-step abstraction hierarchy (see Figure: The FRBR Abstraction Hierarchy.) between the abstract *work*, an *expression* in multiple formats or genres, a particular *manifestation* in one of those formats or genres, and a specific physical *item*. (From “Identity and Bibliographic Resources”.)

See also resource

ITIL

Information Technology Infrastructure Library (ITIL)

(http://www.itil-officialsite.com/)

J

JavaScript Object Notation (JSON)

JavaScript Object Notation (JSON) is a textual format for exchanging data that borrows its metamodel from the JavaScript programming language. Specifically, the JSON metamodel consists of two kinds of structures found in JavaScript: lists (called "arrays" in JavaScript) and dictionaries (called "objects" in JavaScript). (From "JSON".)

(`http://www.json.org/`)

JPEG

Joint Photographic Experts Group

(`http://www.jpeg.org/`)

K

KM

Knowledge Management (KM)

KMS

Knowledge management systems (KMS) are a type of business organizing system whose goal is to capture and systematize these information resources. (From "Preserving Resource Types".)

L

LCC

Library of Congress Classification (LCC)

(http://www.loc.gov/catdir/cpso/lcc.html)

learns

See machine learning.

lexical gap

A **lexical gap** in a language exists when it lacks a word for a concept that is expressed as a word in another language. (From "The Lexical Perspective".)

lexical perspective

The lexical perspective focuses on how the conceptual description of a relationship is expressed using words in a specific language. (From "Describing Relationships: An Overview".)

linguistic relativity

Languages differ a great deal in the words they contain and also in more fundamental ways that they require speakers or writers to attend to details about the world or aspects of experience that another language allows them to ignore. This idea is often described as *linguistic relativity*. (From "Cultural Categories".)

link

See hypertext link

link base

A **link base** is a collection of links stored separately from the resources that they link. (Mentioned in the sidebar, "Perspectives on Hypertext Links".)

link type

When it is evident, this semantic property of the link is called the *link type*. (From "Hypertext Links".)

list

A *list*, like a *set*, is a collection of items with an additional constraint: their items are ordered. (From "Lists".)

literary warrant

The principle of *literary warrant* holds that a classification must be based only on the specific resources that are being classified. (From "Principles Embodied in the Classification Scheme".)

LM

language models (LM)

LMS

Learning Management System (LMS)

loading

Adding resources to a collection.

LOC

Library of Congress (LOC)

(`http://www.loc.gov`)

LOC-CN

Library of Congress Call Number (LOC-CN)

LOC-SH

Library of Congress Subject Headings (LOC-SH)

locative inclusion

Locative inclusion is a type of inclusion relationship between an area and what it surrounds or contains. It is most often expressed using “is-in” as the relationship. However, the entity that is contained or surrounded is not a part of the including one, so this is not a part-whole relationship.

See also “Inclusion”

logical hierarchy

If multiple resource properties are considered in a fixed order, the resulting arrangement forms a *logical hierarchy*. (From “Organizing with Multiple Resource Properties”.)

M

machine learning

machine learning is home to numerous techniques for creating classifiers by training them with already correctly categorized examples. This training is called *supervised learning*; it is supervised because it starts with instances labeled by category, and it involves learning because over time the classifier improves its performance by adjusting the weights for features that distinguish the categories. But strictly speaking, supervised learning techniques do not learn the categories; they implement and apply categories that they inherit or are given to them. (From “Computational Categories”.)

MADS

Metadata Authority Description Standard (MADS)

(`http://www.loc.gov/standards/mads/`)

maintaining

Managing and adapting the resources and the organization imposed on them as needed to support the interactions. (From “Introduction”.)

manifestation

The distinctions put forth by Panizzi, Lubetzky, Svenonius and other library science theorists have evolved today into a four-step abstraction hierarchy (see Figure: The FRBR Abstraction Hierarchy.) between the abstract *work*, an *expression* in multiple formats or genres, a particular *manifestation* in one of those formats or genres, and a specific physical *item*. (From "Identity and Bibliographic Resources".)

MARC

Machine-Readable Cataloging (MARC)

(http://www.loc.gov/marc/)

map

See dictionary

markup

Markup is an encoding of character content with a layer of intentional coding, typically by surrounding the character text with "pointy brackets" or tags whose name suggests a content type, structural role, or formatting. (Ed.)

materiality

It is the requirement to recognize the *materiality* of the environment that enables people to create and interact with digital resources (From "Organizing Digital Resources".)

materials facet

Concerned with the actual substance of which a work is made, like "metal" or "bleach." "Materials" differ from "Physical Attributes" in that the latter is more abstract than the former. (From "Faceted Classification in Description".)

matter facet

One of Ranganathan's universal facets in colon classification. The constituent material of the thing. (From "Foundations for Faceted Classification".)

member-collection inclusion

Member-Collection is the part-whole relationship type where "is-part-of" means "belongs-to," a weaker kind of association than component-object because there is no assumption that the component has a specific role or function in the whole. (From "Types of Semantic Relationships".)

See also inclusion

memory institution

The concept of *memory institution* broadly applies to a great many organizing systems that share the goal of preserving knowledge and cultural heritage. (From "Motivations for Maintaining Resources".)

meronymic inclusion

See part-whole

See also inclusion

MeSH

Medical Subject Headings (MeSH)

(`http://www.nlm.nih.gov/mesh/`)

metadata

Metadata is often defined as "data about data," a definition that is nearly as ubiquitous as it is unhelpful. A more content-full definition of metadata is that it is structured description for information resources of any kind. (From "Metadata".)

See also description resources

metamodels

When common sets of design decisions can be identified that are not specific to any one domain, they often become systematized in textbooks and in design practices, and may eventually be designed into standard formats and architectures for creating organizing systems. These formally recognized sets of design decisions are known as *abstract models* or *metamodels*. *Metamodels* describe structures commonly found in resource descriptions and other information resources, regardless of the specific domain. (From “Structuring Descriptions”.)

metonymy

Part-whole or meronymic semantic relationships have lexical analogues in *metonomy*, when an entity is described by something that is contained in or otherwise part of it. (From “Metonymy”.)

monothetic categories

Monothetic categories are defined by necessary and sufficient properties.

See classical categories

morphemes

See morphology

morphology

The basic building blocks for words are called *morphemes* and can express semantic concepts (when they are called *root words*) or abstract concepts like “pastness” or “plural”). The analysis of the ways by which languages combine *morphemes* is called *morphology*. (From “Relationships among Word Forms”.)

MPAA

Motion Picture Association of America (MPAA)

(http://www.mpaa.org/)

N

n-ary links

n-ary links connect one anchor to multiple types of destinations. (Mentioned in the sidebar, “Perspectives on Hypertext Links”.)

NAICS

North American Industry Classification System (NAICS)

(http://www.census.gov/eos/www/naics/)

name

A *name* is a label for a resource that is used to distinguish one from another. (From “Identity, Identifiers, and Names”.)

name matching

In organizing systems that contain data, there are numerous tools for *name matching*, the task of determining when two different text strings denote the same person, object, or other named entity. (From “Computational Curation”.)

namespace

We can prevent or reduce identifier collisions by adding information about the *namespace*, the domain from which the names or identifiers are selected, thus creating what are often called *qualified names*. (From “Make Identifiers Unique or Qualified”.)

NAPO

National Association of Professional Organizers (NAPO)

(`http://www.napo.net/`)

natural language processing

Natural language processing

navigation

If users are not able to specify their information needs in a way that the *finding* functionality requires, they should be able to use relational and structural descriptions among the resources to navigate from any resource to other ones that might be better. (From “Resource Description to Support Organizing”.)

NCHS

National Center for Health Statistics (NCHS)

(`http://www.cdc.gov/nchs/`)

NCSA

National Center for Supercomputing Applications (NCSA)

(`http://www.ncsa.illinois.edu/`)

NFL

National Football League (NFL)

(http://www.nfl.com/)

NIH

National Institute of Health (NIH)

(http://www.nih.gov/)

NIST

National Institute of Standards and Technology (NIST)

(http://www.nist.gov/)

NLP

Natural Language Processing (NLP)

node

Nodes are objects in an entity-relationship system.

In the RDF metamodel, a pair of nodes and its edge is called a *triple*, because it consists of three parts (two nodes and one edge). The RDF metamodel is a directed graph, so it identifies one node (the one from which the edge is pointing) as the *subject* of the triple, and the other node (the one to which the edge is pointing) as its *object*. The edge is referred to as the *predicate* or (as we have been saying) *property* of the triple. (From "RDF".)

notation

A **notation** a set of characters with distinct forms. (From "Notations".)

The Latin alphabet is a *notation*, as are Arabic numerals. Some more exotic *notations* include alchemical symbols and the symbols used for editorial markup.

NSF

National Science Foundation (NSF)

(http://www.nsf.gov/)

O

OASIS

Organization for the Advancement of Structured Information Standards (OASIS)

(https://www.oasis-open.org/)

object

In the RDF metamodel, a pair of nodes and its edge is called a *triple*, because it consists of three parts (two nodes and one edge). The RDF metamodel is a directed graph, so it identifies one node (the one from which the edge is pointing) as the *subject* of the triple, and the other node (the one to which the edge is pointing) as its *object*. The edge is referred to as the *predicate* or (as we have been saying) *property* of the triple. (From "RDF".)

See also: resource

object warrant

With classifications of physical resources like those in a kitchen, we see *object warrant*, where similar objects are put together, but more frequently the justifying principle will be one of use warrant, where resources are organized based on how they are used. (From “Principles Embodied in the Classification Scheme”.)

objectivity

Although every classification has an explicit or implicit bias (“Classification Is Biased”), facets and facet values should be as unambiguous and concrete as possible to enable reliable classification of instances. (From “Design Principles and Pragmatics”.)

objects facet

The largest facet, *objects* contains the actual works, like “sandcastles” and “screen prints.” (From “Faceted Classification in Description”.)

obtaining

Physical resources often require significant effort to obtain after they have been selected. (From “Resource Description to Support Organizing”.)

OCAD

Ontario Academy of Art and Design (OCAD)

(http://www.ocadu.ca/)

OCLC

Online Computer Library Center (OCLC)

(http://www.oclc.org/)

OECD

Organization for Economic Cooperation and Development (OECD)

(http://www.oecd.org/)

OMG

Object Management Group (OMG)

(http://www.omg.org/)

one-way

Allowing physical or conceptual movement in one direction only. (Ed.)

one-way link

A **one-way link** asserts a link from a resource to one or more resources. A one-way link does not imply a link in the return direction, or among the target resources. (From the Sidebar, “Perspectives on Hypertext Links”.)

See also hypertext

See also directionality

ONIX

Online Information Exchange (ONIX)

(http://www.editeur.org/8/ONIX)

ontology

Ontology is a branch of philosophy concerned with what exists in reality and the general features and relations of whatever that might be. Computer science has adopted *ontology* to refer to any computer-processable resource that represents the relationships among words and meanings in some knowledge domain. (See “Ontologies”)

organize

To *organize* is to create capabilities by intentionally imposing order and structure. (From “The Discipline of Organizing”.)

organizing

Specifying the principles or rules that will be followed to arrange the resources. (From “Introduction”.)

organizing principles

Organizing principles are directives for the design or arrangement of a *collection* of resources that are ideally expressed in a way that does not assume any particular implementation or realization. (From “The Concept of “Organizing Principle””

organizing system

Organizing System: an intentionally arranged collection of resources and the interactions they support. (From “The Discipline of Organizing”.)

orthogonality

Facets should be independent dimensions, so a resource can have values of all of them while only having one value on each of them. (From “Design Principles and Pragmatics”.)

OWL

Web Ontology Language (OWL)

(http://www.w3.org/TR/owl2-overview/)

P

palimpsest

A document or other resource in which the remnants of older content remain visible under the new.

part-whole inclusion

Part-whole inclusion or *meronymic inclusion* is a second type of inclusion relationship. It is usually expressed using “is-part-of,” “is-partly,” or with other similar predicate expressions. (From “Types of Semantic Relationships”.)

See also inclusion

passive resources

Passive resources are usually tangible and static and thus they become valuable only as a result of some action or interaction with them. (From “Passive or Operand Resources”.)

PDF

Portable Document Format (PDF)

(`http://www.adobe.com/products/acrobat/adobepdf.html`)

persistence

Persistence is the quality of resisting change over time. (See “Persistence” and “Identifying Properties”.

personality facet

One of Ranganathan’s universal facets in colon classification.

The type of thing. (From “Foundations for Faceted Classification”.)

phase-activity inclusion

Phase-*Activity* is similar to *feature-activity* except that the phases do not make sense as standalone activities without the context provided by the activity as a whole. (From “Types of Semantic Relationships”.)

See also inclusion

physical attributes facet

Material characteristics that can be measured and perceived, like “height” and “flexibility.” (From “Faceted Classification in Description”.)

PIM

Personal Information Management (PIM)

place-area inclusion

Place-*Area* relationships exist between areas and specific places or locations within them. Like members of collections, places have no particular functional contribution to the whole. (From “Types of Semantic Relationships”.)

See also inclusion

polysemes

If the different meanings of the homographs are related, they are called **polysemes**. (From “Homonymy, Polysemy, and False Cognates”.)

polysemy

Polysemy is the linguistic term for words with multiple meanings or senses. (From “Homonymy, Polysemy, and False Cognates”.)

polythetic

Categories defined by family resemblance or multiple and shifting property sets are termed *polythetic*. (From “Probabilistic Categories and “Family Resemblance””.)

POP

Post Office Protocol (POP)

(`https://tools.ietf.org/html/rfc1939`)

portion-mass inclusion

Portion-Mass is the relationship type when all the parts are similar to each other and to the whole. (From “Types of Semantic Relationships”.)

See also inclusion

possession relationship

Asserting ownership or control of a resource; often expressed using a “has” predicate, such as “has-serial-number-plate.” (From “Types of Semantic Relationships”.)

precision

Precision measures the *accuracy* of a result set, that is, how many of the retrieved resources for a query are relevant. (From “The Recall / Precision Tradeoff”.)

predicate

A *predicate* is a verb phrase template for specifying properties of objects or a relationship among objects. (From “The Semantic Perspective”.)

PREMIS

Preservation Metadata Implementation Strategies (PREMIS) (http://www.loc.gov/standards/premis/)

preservation

Preservation is a maintenance activity.

See also curation, collection development

preservation metadata

Preservation metadata is technical information about resource formats and technology needed to ensure resource and collection integrity in a maintenance context. (From "Resource Description to Support Maintenance".)

primary resource

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

primary subject matter

One of Panofsky's three levels of description for artistic resources. At this level, we describe the most basic elements of a work in a generic way that would be recognizable by anyone regardless of expertise or training. The painting The Last Supper, for example, might be described as "13 people having dinner." (From "Describing Museum and Artistic Resources".)

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." (From "Organizing Resources".)

property-based categorization

Property-based categorization works tautologically well for categories like "prime number" where the category is defined by necessary and sufficient properties. Property-based categorization also works well when properties are conceptually distinct and the value of a property is easy to perceive and examine, as they are with man-made physical resources like shirts. (From "The Limits of Property-Based Categorization".)

propositional synonyms

Propositional synonyms are not identical in meaning, but they are equivalent enough that substituting one for the other will not change the truth value of the sentence. (From "Synonymy".)

provenance

Provenance is the history of the ownership of a collection or the resources in it, where they have been and who has possessed them. In organizing systems like museums and archives that preserve rare or culturally important objects or documents, **provenance** describes a record of who has authenticated a resource over time. (From "Provenance"

Q

QR

Quick Response (QR)

(http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=43655)

qualified names

Qualified names are identifiers which explicitly identify the domain, or namespace, from which they are drawn, thereby reducing identifier collision. (From “Make Identifiers Unique or Qualified”.)

quality

A **quality** is an attribute or property of a resource. A quality is logically ascribable by a subject. (Ed.)

Quality is a measure of the fitness of purpose of a resource or service. It is the difference between what was planned or expected versus what was realized or manifest; it is as an assessment of the suitability of a resource or interaction. (Ed.)

querying

Querying is a very common interaction in many organizing systems, including libraries, museums, archives, computer science, information architecture, data science, the Web, philosophy, cognitive sciences, linguistics, business, and law. Formulating a query in natural language is typically a precursor to application of more systematized techniques discussed throughout this book.

(See especially *Design Decisions in Organizing Systems*, *Interactions with Resources*, and *Case Studies*

R

RDA

Resource Description and Access (RDA)

(http://www.loc.gov/aba/rda/)

RDF

Resource Description Framework (RDF)

(http://www.w3.org/RDF/)

RDF vocabulary

A set of RDF predicate names and URIs is known as an *RDF vocabulary*. (From "Specifying Vocabularies and Schemas".)

reachability

Reachability is the "can you get there from here" property between two resources in a directed graph. (From "Analyzing Link Structures".)

recall

Recall measures the completeness of the result set, that is, how many of the relevant resources in a collection were retrieved. (From “The Recall / Precision Tradeoff”.)

regular expressions

Regular expressions are used in computing for matching text patterns. A regular expression is written in a formal language, which may vary among implementations.

See the sidebar, Regular Expressions in “Controlling Values”.

relationship

A **relationship** is an association among several things, with that association having a particular significance. (From “Introduction”.)

RELAX-NG

REgular LAnguage for XML Next Generation (RELAX NG)

(`http://relaxng.org/`)

relevance

The concept of *relevance* and its relationship to effectiveness is pivotal in information retrieval and machine learning interactions. (From “Effectiveness”.)

reporting

A common interaction with an organizing system.

representation

A principle of good descriptions: Use descriptions that reflect how the resources describe themselves; assume that self-descriptions are accurate. (From “Principles of Good Description”.)

resolution

For a digital resource, its identifier serves as the input to the system or function that determines its location so it can be retrieved, a process called *resolving* the identifier or *resolution*. (From “Identity, Identifiers, and Names”.)

resource

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

Resource has specialized meaning in Internet architecture. It is conventional to describe web pages, images, videos, and so on as *resources*, and the protocol for accessing them, *Hypertext Transfer Protocol* (HTTP), uses the *Uniform Resource Identifier* (URI). (From “The Concept of “Resource””.)

resource description

We describe resources so that we can refer to them, distinguish among them, search for them, manage access to them, preserve them, and make predictions about what might happen to them or what they might do. Each purpose may require different *resource descriptions*. We use *resource descriptions* in every communication and conversation; they are the enablers of organizing systems.

Resource Description Framework (RDF)

The Resource Description Framework (RDF) *metamodel* is a directed graph, so it identifies one node (the one from which the edge is pointing) as the subject of the triple, and the other node (the one to which the edge is pointing) as its object. The edge is referred to as the predicate of the triple. (From “RDF”.)

REST

Representational State Transfer (REST)

RFID

Radio-frequency Identification (RFID)

See US Patent 4,384,288

rich descriptions

Rich descriptions are created by trained and disciplined professionals, often in institutional contexts.

See “Creating Resource Descriptions”

root word

The form of a word after all affixes are removed. (From “Relationships among Word Forms”.)

S

scalability

Facet values must accommodate potential additions to the set of instances. Including an "Other" value is an easy way to ensure that a facet is flexible and hospitable to new instances, but it not desirable if all new instances will be assigned that value. (From "Design Principles and Pragmatics".)

scale

The number of resources and interactions that the collection entails. (Ed.)

schema

A *schema* (or model, or metadata standard) specifies the set of descriptions that apply to an entire resource type. (From "Abstraction in Resource Description".)

scientific warrant

The principle of *scientific warrant* argues that only the categories recognized by the scientists or experts in a domain should be used in a classification system, and it is often opposed by the principle of *use* or *user warrant*, which chooses categories and descriptive terms according to their frequency of use by everyone, not just experts. (From "Principles Embodied in the Classification Scheme".)

scope

The resource *domain* and *scope* circumscribe the describable properties and the possible purposes that descriptions might serve. (From "The Process of Describing Resources".)

secondary subject matter

One of Panofsky's three levels of description for artistic resources. Here, we introduce a level of basic cultural understanding into a description. Someone familiar with a common interpretation of the Bible, for example, could now see The Last Supper as representing Jesus surrounded by his disciples. (From "Describing Museum and Artistic Resources".)

selecting

Determining the scope of the organizing system by specifying which resources should be included.

Selecting in this context means the user activity of using resource descriptions to support a choice of resource from a collection, not the institutional activity of selecting resources for the collection in the first place. (From "Resource Description to Support Organizing".)

self-organizing systems

Self-organizing systems can change their internal structure or their function in response to feedback or changed circumstances. (From "The Concept of "Intentional Arrangement"".)

semantic balance

Top-level facets should be the properties that best differentiate the resources in the classification domain. The values should be of equal semantic scope so that resources are distributed among the subcategories. Subfacets of "Cookware" like "Sauciers and Saucepans" and "Roasters and Brasiers" are semantically balanced as they are both named and grouped by cooking activity. (From "Design Principles and Pragmatics".)

semantic gap

The *semantic gap* is the difference in perspective in naming and description when resources are described by automated processes rather than by people. (From "The Semantic Gap".)

semantic perspective

The **semantic perspective** characterizes the meaning of the association between resources. (From "Describing Relationships: An Overview".)

semantic web

The vision of a *Semantic Web* world builds upon the web world, but adds some further prescriptions and constraints for how to structure descriptions. The Semantic Web world unifies the concept of a resource as it has been developed in this book, with the web notion of a resource as anything with a URI. On the Semantic Web, anything being described must have a URI. Furthermore, the descriptions must be structured as graphs, adhering to the RDF metamodel and relating resources to one another via their URIs. Advocates of Linked Data further prescribe that those descriptions must be made available as representations transferred over HTTP. (From "The Semantic Web World".)

sensemaking

Sensemaking (or sense-making) is the set of processes used by humans to derive meaning from experience or to enhance our understanding. Philosophy, the cognitive sciences and linguistics are among the related academic disciplines.

SEO

Search Engine Optimization (SEO)

set

The simplest way to structure a description is to give it parts and treat them as a *set*. (From “Sets”.)

SGML

Standard Generalized Markup Language (SGML)

(http://www.w3.org/TR/html4/intro/sgmltut.html)

Shepardizing

The analysis of legal citations to determine whether a cited case is still good law is called **Shepardizing** because lists of cases annotated in this way were first published in the late 1800s by Frank Shepard, a salesman for a legal publishing company. (From “Bibliometrics, Shepardizing, Altmetrics, and Social Network Analysis”.)

SKOS

Simple Knowledge Organization System (SKOS)

(http://www.w3.org/2004/02/skos/)

SKU

Stock Keeping Unit (SKU)

similarity

Similarity is a measure of the resemblance between two things that share some characteristics but are not identical. It is a very flexible notion whose meaning depends on the domain within which we apply it. (From “Similarity”.)

smart things

See active resources.

social classification

Using any property of a resource to create a description is an uncontrolled and often unprincipled principle for creating categories is called **social classification** or **tagging**. (From "Classification vs. Tagging".)

SNOMED-CT

Systematized Nomenclature of Medicine — Clinical Terms (SNOMED-CT)

(`http://www.ihtsdo.org/snomed-ct/`)

SOA

Service Oriented Architecture

space facet

One of Ranganathan's universal facets in colon classification. Where the thing occurs. (From "Foundations for Faceted Classification".)

spectrum facets

Assume a range of numerical values with a defined minimum and maximum. Price and date are common spectrum facets. The ranges are often modeled as mutually exclusive regions (potential price facet values might include "$0—$49," "$50—$99," and "$100—$149"). (From "A Classification for Facets".)

SQL

Structured Query Language (SQL)

ISO/IEC 9075:2011 "Information technology – Database languages – SQL"

standardization

A principle of good description: Standardize descriptions to the extent practical, but also use aliasing to allow for commonly used terms. (From "Principles of Good Description".)

statistical pattern recognition

See unsupervised learning

stemming

These processing steps normalize inflectional and derivational variations in terms, e.g., by removing the "-ed" from verbs in the past tense. This homogenization can be done by following rules (*stemming*) or by using dictionaries (*lemmatization*). Rule-based stemming algorithms are easy to implement, but can result in wrongly normalized word groups, for example when "university" and "universe" are both stemmed to "univers." (From "Transforming Resources for Interactions".)

stopword elimination

Stopwords are those words in a language that occur very frequently and are not very semantically expressive. Stopwords are usually articles, pronouns, prepositions, or conjunctions. Since they occur in every text, they can be removed because they cannot distinguish them. Of course, in some cases, removing stopwords might remove semantically important phrases (e.g., "To be or not to be"). (From "Transforming Resources for Interactions".)

storage

Storage is a maintenance activity.

See also preservation, curation

structural perspective

The **structural perspective** analyzes the patterns of association, arrangement, proximity, or connection between resources without primary concern for their meaning or the origin of these relationships. (From "Describing Relationships: An Overview".)

structured descriptions

See "Creating Resource Descriptions"

stuff-object inclusion

Stuff-Object relationships are most often expressed using "is-partly" or "is-made-of" and are distinguishable from component-object ones because the stuff cannot be separated from the object without altering its identity. The stuff is not a separate ingredient that is used to make the object; it is a constituent of it once it is made. (From "Types of Semantic Relationships".)

See also inclusion

styles and periods facet

Artistic and architectural eras and stylistic groupings, such as "Renaissance" and "Dada." (From "Faceted Classification in Description".)

subject

In the RDF metamodel, a pair of nodes and its edge is called a *triple*, because it consists of three parts (two nodes and one edge). The RDF metamodel is a directed graph, so it identifies one node (the one from which the edge is pointing) as the *subject* of the triple, and the other node (the one to which the edge is pointing) as its *object*. The edge is referred to as the *predicate* or (as we have been saying) *property* of the triple. (From "RDF".)

sufficiency and necessity

Descriptions should have enough information to serve their purposes and not contain information that is not necessary for some purpose; this might imply excluding some aspects of self-descriptions that are insignificant. (From "Principles of Good Description".)

supervised learning

In *supervised learning*, a machine learning program is trained with sample items or documents that are labeled by category, and the program learns to assign new items to the correct categories. (From "Computational Categories")

surrogate resource

See description resources.

SUV

Sport Utility Vehicle (SUV)

SVM

Support Vector Machine (SVM)

symmetric relationships

Symmetric relationships are bi-directional; they express the same relationship from the subject to object as they do from the object to the subject. For example, "is-married-to."

synonym

When something has more than one name, each of the multiple names is a **synonym** or **alias**. (From "Synonymy".)

synonymy

Synonymy is the relationship between words that express the same semantic concept. (From "Synonymy".)

synset

An unordered set of synonyms is often called a **synset**. Synsets are interconnected by both semantic relationships and lexical ones, enabling navigation in either space. (From "Synonymy".)

syntax

The *syntax* and *grammar* of a language consists of the rules that determine which combinations of its words are allowed and are thus grammatical or *well-formed*. Natural languages have substantial similarities by having nouns, verbs, adjectives and other parts of speech, but they differ greatly in how they arrange them to create sentences. (From "Syntax and Grammar".)

T

tag cloud

Folksonomies are often displayed in the form of a **tag cloud**,where the frequency with which the tag is used throughout the site determines the size of the text in the tag cloud. The tag cloud emerges through the bottom-up aggregation of user tags and is a statistical construct, rather than a semantic one. (From “Classification vs. Tagging”

tagging

Using any property of a resource to create a description is an uncontrolled and often unprincipled principle for creating categories is called **social classification** or **tagging**. (From “Classification vs. Tagging”

tagsonomy

When users or communities establish sets of principles to govern their tagging practices, tagging is even more like classification. Such a tagging system can be called a *tagsonomy*, a neologism we have invented to describe more systematic tagging. (From “Classification vs. Tagging”

taskonomy

A task or activity-based classification system is called a **taskonomy**. (From “Classification by Activity Structure”

taxonomic classification

When multiple resource properties are considered in a fixed sequence, each property creates another level in the system of categories and the classification scheme is **hierarchical** or **taxonomic**. (From “Classification by Activity Structure”

taxonomic facets

Taxonomic facets, also known as hierarchical facets are based on logical containment. (From “A Classification for Facets”

taxonomy

A **taxonomy** is a hierarchy that is created by a set of interconnected class inclusion relationships. (From “Inclusion”

See also inclusion

TCP/IP

Transmission Control Protocol/Internet Protocol (TCP/IP)

(https://tools.ietf.org/html/rfc1180)

TEI

Text Encoding Initiative (TEI)

(http://www.tei-c.org/index.xml)

temporal inclusion

Temporal inclusion is a type of inclusion relationship between a temporal duration and what it surrounds or contains. It is most often expressed using "is-in" as the relationship. However, the entity that is contained or surrounded is not a part of the including one, so this is not a part-whole relationship. (From "Inclusion".)

See also inclusion

term frequency

A vector space ranking utilizes an intrinsic resource property, the number of individual terms in a resource, called the **term frequency**. For each term, term frequency measures how many times the term appears in a resource. (From "Ranked Retrieval with Vector Space or Probabilistic Models"

theory-based category

A final psychological principle for creating categories is organizing things in ways that fit a theory or story that makes a particular categorization sensible. A *theory-based category* can win out even if probabilistic categorization, on the basis of *family resemblance* or *similarity* with respect to visible properties, would lead to a different category assignment. (From “Theory-Based Categories”.)

thesaurus

A **thesaurus** is a reference work that organizes words according to their semantic and lexical relationships. Thesauri are often used by professionals when they describe resources. (From “Thesauri”.)

time facet

One of Ranganathan’s universal facets in colon classification. When the thing occurs. (From “Foundations for Faceted Classification”.)

tokenization

Segments the stream of characters (in an encoding scheme, a space is also a character) into textual components, usually words. In English, a simple rule-based system can separate words using spaces. However, punctuation makes things more complicated. For example, periods at the end of sentences should be removed, but periods in numbers should not. Other languages introduce other problems for tokenization; in Chinese, a space does not mark the divisions between individual concepts. (From the sidebar Text Processing in “Transforming Resources for Interactions”.)

topological inclusion

Topological inclusion is a type of inclusion relationship between a container and what it surrounds or contains. It is most often expressed using “is-in” as the relationship. However, the entity that is contained or surrounded is not a part of the including one, so this is not a part-whole relationship. (From “Inclusion”.)

See also inclusion

training set

A **training set** for supervised learning is taken from the labeled instances. The remaining instances are used for validation. (From "Computational Classification".)

transclusion

The inclusion, by hypertext reference, of a resource or part of a resource into another resource is called *transclusion*. Transclusion is normally performed automatically, without user intervention. The inclusion of images in web documents is an example of transclusion. Transclusion is a frequently used technique in business and legal document processing, where re-use of consistent and up-to-date content is essential to achieve efficiency and consistency. (From "Hypertext Links")

transformation

Transformation is a very broad concept but in the context of organizing systems it typically means a change in a resource representation or description. The transformation can involve the selection, restructuring, or rearrangement of resources or parts of them. (See "Transforming Resources for Interactions".)

transitivity

Transitivity is another property that can apply to semantic relationships. When a relationship is transitive, if X and Y have a relationship, and Y and Z have the same relationship, then X also has the relationship with Z. Any relationship based on ordering is transitive, which includes numerical, alphabetic, and chronological ones as well as those that imply qualitative or quantitative measurement. (From "Transitivity".)

tree

Trees consist of nodes joined by edges, recursively nested. When a single, root dictionary is connected to child nodes that are themselves dictionaries, we say that the dictionaries are *nested* into a kind of **tree** structure.

A *tree* is a constrained *graph*. Trees are *directed* graphs because the “parent of” relationship between nodes is asymmetric: the edges are arrows that point in a certain direction. Trees are *acyclic* graphs, because if you follow the directed edges from one node to another, you can never encounter the same node twice. Finally, trees have the constraint that every node (except the root) must have exactly one parent. (From “Trees”.)

triple

In the RDF metamodel, a pair of nodes and its edge is called a *triple*, because it consists of three parts (two nodes and one edge). The RDF metamodel is a directed graph, so it identifies one node (the one from which the edge is pointing) as the *subject* of the triple, and the other node (the one to which the edge is pointing) as its *object*. The edge is referred to as the *predicate* or (as we have been saying) *property* of the triple. (From “RDF”.)

TXL

Turing eXtender Language (TXL)

(`http://www.txl.ca/`)

typicality

Typicality or **centrality** considers some members of the category better examples than others, even if they share most properties. (From “Probabilistic Categories and “Family Resemblance””.)

U

UBL

Universal Business Language (UBL)

(https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=ubl)

UDC

Universal Decimal Classification (UDC)

(http://www.udcc.org/)

UK

United Kingdom (UK)

(https://www.gov.uk/)

UN

United Nations (UN)

(http://www.un.org/en/)

uniqueness principle

The *uniqueness principle* means the categories in a classification scheme are mutually exclusive. Thus, when a logical concept is assigned to a particular category, it cannot simultaneously be assigned to another category. (From "Principles for Assigning Resources to Categories".)

UNSPC

United Nations Standard Products and Services Code (UNSPC)

(http://www.unspsc.org/)

unsupervised learning

In *unsupervised learning*, the program gets the same items but has to come up with the categories on its own by discovering the underlying correlations between the items; that is why unsupervised learning is sometimes called *statistical pattern recognition*. (From: "Computational Categories")

See also: machine learning and supervised learning

URI

Uniform Resource Identifier (URI)

(http://www.w3.org/Addressing/)

URL

Uniform Resource Locator (URL)

(http://www.w3.org/TR/url/)

URN

Uniform Resource Name (URN)

(http://www.w3.org/TR/uri-clarification/)

user convenience

Choose description terms with the user in mind; these are likely to be terms in common usage among the target audience. (From "Principles of Good Description".)

user warrant

The principle of *scientific warrant* argues that only the categories recognized by the scientists or experts in a domain should be used in a classification system, and it is often opposed by the principle of *use* or *user warrant*, which chooses categories and descriptive terms according to their frequency of use by everyone, not just experts. (From "Principles Embodied in the Classification Scheme".)

UUID

Universally Unique Identifier (UUID)

(`http://www.ietf.org/rfc/rfc4122.txt`)

V

validation

Validation is the process of verifying that a document or data structure conforms with its schema or schemas. Markup validation confirms the structure of the document. Type validation confirms that the content of leaf nodes conforms with the specification of data types. Content validation confirms that the values of the leaf nodes are appropriate. Link validation confirms the integrity of the links between nodes and between documents. Cross validation is the method commonly used for model selection. Business rule validation confirms compliance with business rules. (Discussed in "Implementing Categories Defined by Properties", "Design Principles and Pragmatics", "Specifying Vocabularies and Schemas"

value

We distinguish between the type of the *attribute* and the *value* that it has. For example, the color of any object is an *attribute* of the object, and the *value* of that attribute might be "green." (From "Attribution".)

VIAF

Virtual International Authority File (VIAF)

(`http://viaf.org/`)

viewing

Viewing is a central interaction in museums and zoos.

See also collection development

VIN

Vehicle Identification Number (VIN)

(ISO 3779:2009)

visiting

Visiting is a central interaction in museums and zoos.

See also collection development

visualization

A common interaction with an organizing system.

vocabulary problem

Every natural language offers more than one way to express any thought, and in particular there are usually many words that can be used to refer to the same thing or concept. (From "The Vocabulary Problem".)

VPN

Virtual Private Network

W

W3C

World Wide Web Consortium (W3C)

(http://www.w3.org/)

warrant principle

The *warrant* principle concerns the justification for the choice of categories and the names given to them. (From “Principles Embodied in the Classification Scheme”.)

See also: literary warrant, scientific warrant, user warrant and object warrant

well-formed

The *syntax* and *grammar* of a language consists of the rules that determine which combinations of its words are allowed and are thus grammatical or *well-formed*. Natural languages have substantial similarities by having nouns, verbs, adjectives and other parts of speech, but they differ greatly in how they arrange them to create sentences. (From “Syntax and Grammar”.)

WHO

World Health Organization (WHO)

(`http://www.who.int/en/`)

work

An abstract idea of an author's intellectual or artistic creation.

The distinctions put forth by Panizzi, Lubetzky, Svenonius and other library science theorists have evolved today into a four-step abstraction hierarchy (see Figure: The FRBR Abstraction Hierarchy.) between the abstract *work*, an *expression* in multiple formats or genres, a particular *manifestation* in one of those formats or genres, and a specific physical *item*.

writing system

A *writing system* employs one or more notations, and adds a set of rules for using them. Most writing systems assume knowledge of a particular human language. These writing systems are known as *glottic* writing systems. But there are many writing systems, such as mathematical and musical ones, that are not tied to human languages in this way. Many of the writing systems used for describing resources belong to this latter group, meaning that (at least in principle) they can be used with equal facility by speakers of any language. (From "Writing Systems".)

Some writing systems, such as XML and JSON, are closely identified with specific metamodels.

WSDL

Web Services Description Language (WSDL)

(`http://www.w3.org/TR/wsdl`)

X

XCBF

XML Common Biometric Format (XCBF)

(https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xcbf)

XInclude

XML Inclusions (XInclude)

(http://www.w3.org/TR/xinclude/)

XML

Extensible Markup Language (XML)

(http://www.w3.org/XML/)

XML *Information Set*

The XML *Infoset* is a tree structure, where each node of the tree is defined to be an "information item" of a particular type. Each information item has a set of type-specific properties associated with it. At the root of the tree is a "document item," which has exactly one "element item" as its child. An *element* item has a set of *attribute* items, and a list of child nodes. These child nodes may include other element items, or they may be character items. (See "Kinds of Structures" below for more on characters.) *Attribute* items may contain character items, or they may contain typed data, such as name tokens, identifiers and references. Element identifiers and references (ID/IDREF) may be used to connect nodes, transforming a tree into a graph. (From "XML Information Set".)

XSD

XML Schema Definition Language (XSD)

(http://www.w3.org/XML/Schema.html)

XSLT

Extensible Stylesheet Language Transformations (XSLT)

Based on XML, XSLT is a declarative language designed for transforming XML documents into other documents. For example, XSLT can be used to convert XML data into HTML documents for web display or PDF for print or screen display. XSLT processing entails taking an input document in XML format and one or more XSLT style sheets through a template-processing engine to produce a new document.

(http://www.w3.org/TR/xslt)

Z

zoo

A **zoo** is an organizing system for living animals that arranges them according to principles of biological taxonomy or common habitat. (Ed.)

Index

MURRAY MALONEY

An Appendix to The Discipline of Organizing

A

B

C

concept

E

F

G

H

I

J

J

K

K

L

M

N

O

P

Q

R

T

U

X

Y

Z

Colophon

MURRAY MALONEY

An Appendix to The Discipline of Organizing

An Intentional Arrangement

Progressive intentional arrangements guided by organizing principles.

The cover is An Intentional Arrangement by Jen Wang. The cover font is HTF Knockout.

The authors created textual content using various versions of Microsoft Word. The editors managed, organized and massaged the content in Word files using the DropBox cloud. Those Word documents were then sent out for transformation into DocBook representations using scripts and hand editing. Then the DocBook files were encoded with extensive semantic tagging and hypertext relationships, including bibliographic and index references, and transclusion of content from the chapters into the quizzes and glossary. The editors used vi, oXygen XML Editor, XML Spy and XMetaL word processors on OS X and Windows platforms.

The content of the printed book, as well as the third and fourth edition PDF, is presented using OpenType DejaVu fonts. The text font is DejaVuSerif; the heading font is DejaVuSans. The monospaced font used for code samples and literal text is DejaVuMono. The condensed font used in the index is DejaVuSerifCondensed; some labels use DejaVuSansCondensed. A small number of characters require a Unicode font to represent them; the Unicode font used for print and PDF versions of this book is Adobe Song Standard. Emphasis is presented in bold, italic, or oblique variants of the corresponding font families.

The PDF manifestation of the third and fourth editions of *The Discipline of Organizing* were built in oXygen using customized XSLT scripts to profile and transform the DocBook5 files into XHTML5 files, which are then processed using CSS and the Antenna House Formatter V6 to produce an accessible paginated Portable Document Format document. The first printed edition of *The Discipline of Organizing* was built with Altas from O'Reilly Media.

The ePub3 electronic book editions offered by O'Reilly Media are built from the DocBook 5 sources using customized XSLT scripts created by Bob Stayton, Sagehill Enterprises and maintained by Jirka Kosek. The scripts take advantage of discipline and user-level values encoded in DocBook attributes to produce profiled, static ePub 3 editions.

The Pressbooks editions were built from the DocBook 5 sources using customized XSLT scripts created by Bob Stayton, Sagehill Enterprises.