

Cataloging Theory in Search of Graph Theory and Other Ivory Towers

Ronald J. Murray and
Barbara B. Tillett

Object: Cultural Heritage Resource Description Networks

This paper summarizes a research program that focuses on how catalogers, other cultural heritage information workers, web/Semantic Web technologists, and the general public understand, explain, and manage resource description tasks by creating, counting, measuring, classifying, and otherwise arranging descriptions of cultural heritage resources within the bibliographic universe and beyond it. A significant effort is made to update the nineteenth-century mathematical and scientific ideas present in traditional cataloging theory to their twentieth- and twenty-first-century counterparts. There are two key elements in this approach: (1) a technique for diagrammatically depicting and manipulating large quantities of individual and grouped bibliographic entities and the relationships between them, and (2) the creation of resource description exemplars (problem–solution sets) that are intended to play theoretical, pedagogical, and IT system design roles.

To The Reader: This paper presents a major re-visioning of cataloging theory, introducing along the way a technique for depicting diagrammatically large quantities of bibliographic entities and the relationships between them. As many details of the diagrams cannot be reproduced in regularly sized print publications, the reader is invited to follow the links provided in the endnotes to PDF versions of the figures.

Cataloging—the systematic arrangement of resources through their descriptions that is practiced by libraries, archives, and museums (i.e., cultural heritage institutions) and other parties¹—can be placed in an advanced, twenty-first-century context by updating its preexisting scientific and mathematical ideas with their more contemporary versions. Rather than directing our attention to implementation-oriented details such as metadata formats, database designs, and communications protocols, as do technologists pursuing bottom-up web and Semantic Web initiatives, in

this paper we will define a complementary, top-down approach.

This top-down approach focuses on how catalogers, other cultural heritage information workers, web/Semantic Web technologists, and the general public have understood, explained, and managed their resource description tasks by creating, counting, measuring, classifying, and otherwise arranging descriptions of cultural heritage resources within and beyond the bibliographic universe. We go on to prescribe what enlargements of cataloging theory and practice are required such that catalogers and other interested parties can describe pages from unique, ancient codices as readily as they might describe information elements and patterns on the web.

We will be enhancing cataloging theory with concepts from communications theory, history of science, graph theory, computer science, and from the hybrid field of anthropology and mathematics called *ethnomathematics*. Employing this strategy benefits two groups:

- Workers in the cultural heritage realm, who will acquire a broadened perspective on their resource description activities, who will be better prepared to handle new forms of creative expressions as they appear, and who will be able to shape the development of information systems that support more sophisticated types of resource descriptions and ways of exploring those descriptions. To build a better library system (perhaps an n-dimensional, n-connected system?), one needs better theories about the library collections and the people or groups who manage and use them.
- The full spectrum of people who draw on cultural heritage resources: scholars, creatives (novelists, poets, visual artists, musicians, and so on), professional and technical workers, students, and other people or groups pursuing specific or general, long- or short-term interests, entertainment, etc.

To apply a multidisciplinary perspective to the processes by which resource description data (linked or otherwise) are created and used is not an ivory tower exercise. Our approach draws lessons from the debates on *why*, *what*, and *how* to describe physical phenomena that were conducted by physicists, engineers, software developers (and their historian and philosopher of science observers) during the evolution of high-energy physics. During that time, intensive debates raged over theory and observational/experimental data, the roles of theorists, experimenters, and instrument builders, instrumentation, and hardware/software system design.² Accommodating the resulting scientific approaches to description, collaboration, and publishing has required the creation of information technologies that have had and continue to have world-shaking effects.

Ronald J. Murray (rmur@loc.gov) is a Digital Conversion Specialist in the Preservation Reformatting Division, and **Barbara B. Tillett** (btill@loc.gov) is the Chief of the Policy and Standards Division at the Library of Congress.

These physics research facilities and their supporting academic institutions are the same ones whose scientific subcultures (theory, experiment, and instrument-building) generated the data creation, management, analysis, and publication requirements that resulted in the creation of the web. In response to this development, we have come to believe that cultural heritage resource description (i.e., the process of identifying and describing phenomena in the bibliographic universe as opposed to the physical one) must now be as open to the concepts and practices of those twenty-first-century physics subcultures as it had been to the natural sciences during the nineteenth century.³ We have consequently undertaken an intensive study of the scientific subcultures that generate scientific data and have identified four principles on which to base a more general approach to cultural heritage resource description:

1. observations
2. complementarity
3. graphs
4. exemplars

The cultural heritage resource description theory to follow proposes a more articulated view of the complex, collaborative process of making available—through their descriptions—socially relevant cultural heritage resources at a global scale. We will demonstrate that a broader understanding of this resource description process (along with the ability to create improved implementations of it) *requires* integrating ideas from other fields of study, reaching beyond IT system design to embrace larger issues.

Cataloging as Observation

As stated in the *Oxford English Dictionary*, an observation is:

The action or an act of observing scientifically; esp. the careful watching and noting of an object or phenomenon in regard to its cause or effect, or of objects or phenomena in regard to their mutual relations (contrasted with experiment). Also: a measurement or other piece of information so obtained; an experimental result.⁴

Following the scientific community's lead in striving to describe the physical universe through observations, we adapted the concept of an observation into the bibliographic universe and assert that cataloging is a process of making observations on resources. Human or computational observers following institutional business rules (i.e., the terms, facts, definitions, and action assertions that represent constraints on an enterprise and on the things of interest to the enterprise)⁵ create *resource*

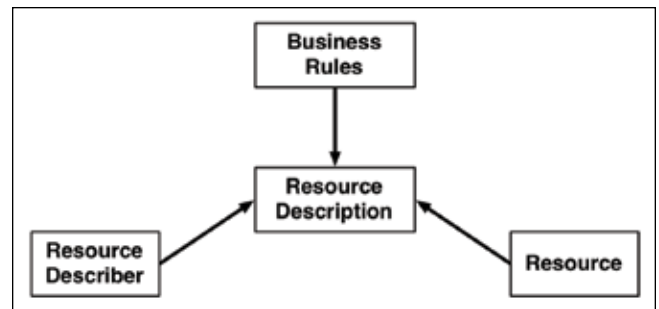


Figure 1. A Resource Description Modeled as a Business Rule-Constrained Account of a Person, Object, or Event

descriptions—accounts or representations of a person, object, or event being drawn on by a person, group, institution, and so on, in pursuit of its interests.

Given this definition, a person (or a computation) operating from a business rules-generated institutional or personal point of view, and executing specified procedures (or algorithms) to do so, is an integral component of a *resource description process* (see figure 1). This process involves identifying a resource's textual, graphical, acoustic, or other features and then classifying, making quality and fitness for purpose judgments, etc., on the resource. Knowing which institutional or individual points of view are being employed is essential when parties possessing multiple views on those resources describe cultural heritage resources. How multiple resource descriptions derived from multiple points of view are to be related to one another becomes a key theoretical issue with significant practical consequences.

Niels Bohr's Complementarity Principle and the Library

In 1927, the physicist Niels Bohr offered a radical explanation for seemingly contradictory observations of physical phenomena confounding physicists at that time.⁶ According to Bohr, creating *descriptions* of nature is the primary task of the physicist:

It is wrong to think that the task of physics is to find out how nature is. Physics concerns what we can say about nature.⁷

Descriptions that appear contradictory or incomparable may in fact be signaling deep limitations in language. Bohr's complementarity principle states that a complete description of atomic-level phenomena requires descriptions of both wave and particle properties. This is generally understood to mean that in the normal language

that physicists use to communicate experimental results, the wholeness of nature is accessible only through the embrace of complementary, contradictory, and paradoxical descriptions of it. Later in his career, Bohr vigorously affirmed his belief that the complementarity principle was not limited to quantum physics:

In general philosophical perspective, it is significant that, as regards analysis and synthesis in other fields of knowledge, we are confronted with situations reminding us of the situation in quantum physics. Thus, the integrity of living organisms, and the characteristics of conscious individuals, and most of human cultures, present features of wholeness, the account of which implies a typically complementary mode of description. . . . We are not dealing with more or less vague analogies, but with clear examples of logical relations which, in different contexts, are met with in wider fields.⁸

Within a library, there are many things catalogers, conservators, and preservation scientists—each with their distinctive skills, points of view, and business rules—can observe and say about cultural heritage resources.⁹ Much of what these specialists say and do strongly affects library users' ability to discover, access, and use library resources in their original or surrogate forms. While observations made by these specialists from different perspectives may lead to descriptions that must be accepted as valid for those specialists, a fuller appreciation of these descriptions calls for the integration of those multiple perspectives into a well-articulated, accessible whole.

Reflecting the perspectives of the Library of Congress directorates in which we work, the Acquisitions and Bibliographic Access (ABA) Directorate and the Preservation Directorate, we assert that the most fundamental complementary views on cultural heritage resources involve describing a library's resources in terms of their availability (from an acquisitions perspective), in terms of their information content (from a cataloging perspective), and in terms of their physical properties (from a preservation perspective). For example, in the normal languages used to communicate their results, Preservation Directorate conservators narrate their condition assessments and record simple physical measurements of library-managed objects—while at the same time preservation scientists in another section bring instrumentation to acquire optical and chemical data from submitted materials and from reference collections of physical and digital media. Even though these assessments and measurements may not be comprehended by or made accessible to most library users, the information gathered possess a critical logical relationship to bibliographic and other descriptions of those same resources. Key decisions regarding a library resource's fitness for

purpose, its reformatting, and its long-term preservation must take into consideration that resource's physical characteristics.

Having things to say about cultural heritage resources—and having many “voices” with which to say them—presents the problem of creating a well-articulated context for library-generated resource descriptions as well as those from other sources. These contextualization issues must be addressed theoretically *before* implementation-level thinking, and the demands of contextualization require visualization tools to complement the narratives common to catalogers, scholars, and other users. This is where mathematics and ethnomathematics make their entrance.

Ethnomathematics is the study of the mathematical practices of specific cultural groups over the course of their daily lives and as they deal with familiar and novel problems.¹⁰ An ethnomathematical perspective on cultural heritage resource description directs one's attention to the existence of simple and complex resource descriptions, the patterns of descriptions that have been created, and the representation of these patterns when they are interpreted as expressions of mathematical ideas. A key advantage of operating from an ethnomathematical perspective is becoming aware that mathematical ideas can be observed within a culture (namely the people and institutions who play key roles in observing the bibliographic universe) before their having been identified and treated formally by Western-style mathematicians.

Resource Description as Graph Creation

Relationships between cultural heritage resource descriptions can be represented as conceptually engaging and flexible systems of connections mathematicians call graphs.

A full appreciation of two key mathematical ideas underlying the evolution of cataloging—putting things into groups and defining relationships between things and groups of things—was only possible after the founding, naming, and expansion of graph theory, which is a field of mathematics that emerged in the 1850s, and the eventual acceptance around 1900 of set theory, a field founded amid intense controversy in 1874. Between the emergence of formal mathematical treatments of those ideas by mathematicians and their actual exploitation by cataloging theorists—or by anyone capable of considering library resource description and organization problems from a mathematical perspective—lay a gulf of more than one hundred years.¹¹ It remained for scholars in the library world to begin addressing the issue. Tillett's 1987 work on bibliographic relationships and Svenonius's 2000 definition of bibliographic entities in set-theoretic terms

identified those mathematical ideas in cataloging theory and developed them formally.¹² Then in 2009, we were able to employ graph theory (expressed in set-theoretical terms *and* in its highly informative graphical representation) as part of a broader historical and cultural analysis.¹³

Cataloging theory had by 2009 haltingly embraced a new view on how resources in libraries have been described and arranged via their descriptions—an activity that in principle stretches back to catalogs created for the Library of Alexandria¹⁴—and how these structured resource descriptions have evolved over time, irrespective of implementation. Murray’s investigation into this issue revealed that the increasingly formalized and refined rules that guided Anglo-American catalogers had, by 1876, specified sophisticated systems of cross-references (i.e., connections between bibliographic descriptions of works, authors, and subjects)—systems whose properties were not yet the subject of formal mathematical treatment by mathematicians of the time.¹⁵ Murray also found that library resource description structures—when teased out of their book and card and digital catalog implementations and treated as graphs—are arguably more sophisticated than those being explored in the World Wide Web Consortium’s (W3C) Library Linked Data initiative.¹⁶

Implementation-Oriented Substitutes For Graph Theory

Cataloging theory has been both helped and hindered by the use of information technology (IT) techniques like entity-relationship modeling (E-R, first used extensively by Tillett in 1987 to identify bibliographic relationships in cataloging records) and object-oriented (OO) modeling.¹⁷ E-R and OO modeling may be used effectively to create information systems that are based on an inventory of “things of interest” and the relationships that exist between them. Unfortunately, the things of interest in cultural heritage institutions keep changing and may require redefinition, aggregation, disaggregation, and re-aggregation. E-R and OO modeling as usually practiced are not designed to manage the degree and kind of changes that take place under those circumstances.

When trying to figure out what is “out there” in the bibliographic universe, we assert that focus should first be placed on identifying and describing the things of interest, what relationships exist between them, and what processes are involved in the creation, etc., of resource descriptions. Having accomplished this, attention can then be safely paid to defining and managing information deemed essential to the enterprise, that is, undertaking IT system analysis and design. But when an IT-centric modeling technique becomes the bed on which the resource description theory itself is constructed, the resulting theory will be driven in a direction that is strongly influenced

by the modeling technique. What is required instead is theory-based guidance of systems development, alongside theory testing and improvement through application use. If software development is not constrained by a tacit or explicit resource description theory or practice, graph or other data structures familiar to the historically less well-informed, those favored by an institution’s system designers and developers, or those familiar to and favored by implementation-oriented communities may be invoked inappropriately.¹⁸

Given graph theory’s potentially overwhelming mathematical power—as evidenced by its many applications in the physical sciences, engineering, and computer science—investigations into graph theory and its history require close attention both to the history and evolving needs of the cultural heritage community.¹⁹ The unnecessary constraint on resource description theory formation occasioned by the use of E-R or OO modeling can be removed by dispensing with IT system analysis tools and expressing resource description concepts in graph-theoretical terms. With this step, the very general elements (i.e., entities and relationships) that characterize E-R models and the more implementation-oriented ones in OO models are replaced by more mathematically flexible, theory-relevant elements expressed in graph-theoretical terms. The result is a “graph-friendly” theory of cultural heritage resource description, which can borrow from other fields (e.g., ethnomathematics, history of science) to improve its descriptive and predictive power, guide IT system design and use, and, in response to users’ experiences with functioning systems, results in improved theories and information systems.

Graph Theory in a Cultural Heritage Context

Ever since the nineteenth century foundation of graph theory (though scholars regularly date its origins from Euler’s 1736 paper)²⁰ and its move from the backwaters of recreational mathematics to full field status by 1936, graph theory has concerned itself with the properties of systems of connections—nowadays regularly expressed as the mathematical objects called sets.²¹ In addition to its set notational form, graphs also are depicted and manipulated in diagrammatic form as dots/labeled nodes linked by labeled or unlabeled, simple or arrowed lines. For example, the graph *X*, consisting of one set of nodes labeled *A*, *B*, *C*, *D*, *E*, and *F* and one set of edges labeled *AB*, *BD*, *DE*, *EF*, and *FC*, can be depicted in set notation as $X = \{\{A\ B\ C\ D\ E\ F\}, \{AB\ BD\ DE\ EF\ FC\}\}$ and can be depicted diagrammatically as in figure 2.

When graphs are defined to represent different types of nodes and relationships, it becomes possible to create and discuss structures that can support cultural heritage resource description theory and application building. The following diagrams depict simple resource description

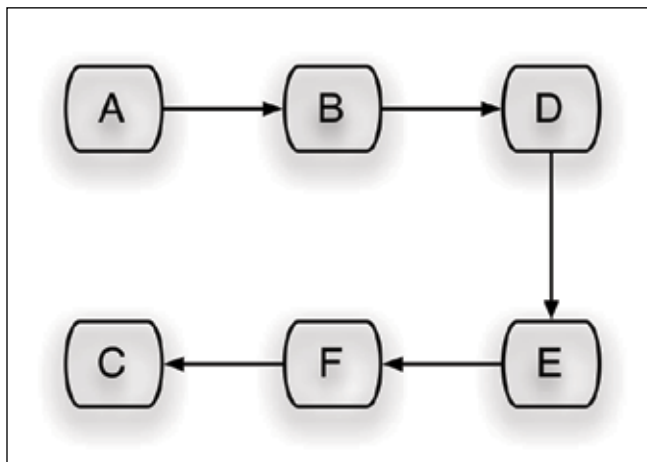


Figure 2. A Diagrammatic Representation of Graph X

graphs that are based on real-world bibliographic descriptions. Nodes in the graphs represent text, numbers, or dates and relationships that can be nondirectional (as a simple line), unidirectional (as single arrowed lines) or bidirectional (as a double arrowed line).

The all-in-one resource description graph in figure 3 can be divided and connected according to the kinds of relationships that have been defined for cultural heritage resources. This is the point where institutional, group, and individual ways of describing resources shape the initial structure of the graph. Once constructed, graph structures like this and their diagrammatic representations are then interpreted in terms of a tacit or explicit resource description theory. In the case of graphs constructed according to IFLA's Functional Requirements for Bibliographic Records (FRBR) standard,²² figure 3 can be subdivided into four FRBR sub-graphs, yielding figure 4.

The four diagrams depict the initial graph of cataloging data as four *complementary* FRBR WEMI (W–work, E–expression, M–manifestation, and I–item) graphs. Note that the item graph contains the call numbers (used here to identify the location of the copy) of three physical copies of the novel. This use of call numbers is qualitatively different from the values found in the manifestation graph in that resource descriptions in this graph apply to the entire population of physical copies printed by the publisher.

The descriptions contained in figure 4's FRBR sub-graphs reproduce bibliographic characteristics found useful by catalogers, scholars, other educationally oriented end users, and to varying extents the public in general. Once created, resource description graphs and subgraphs (in mathematical notation or in simple diagrams like figure 4) can proliferate and link in multiple and complex ways—in parallel with or independently

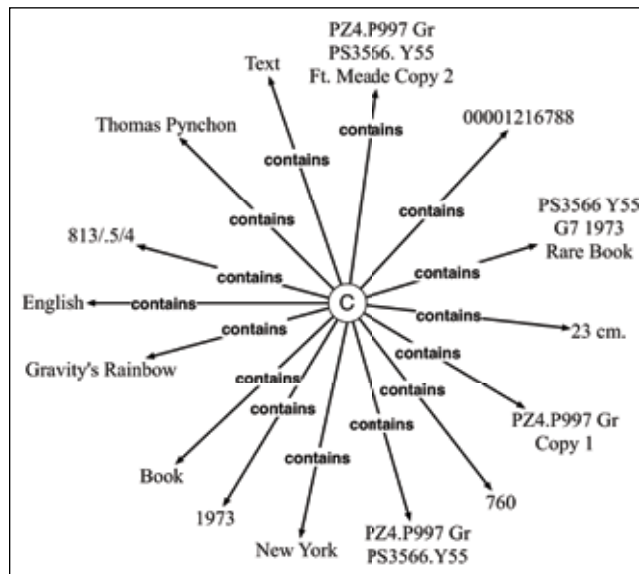


Figure 3. Library of Congress Catalog Data for Thomas Pynchon's Novel *Gravity's Rainbow*, Represented as an All-In-One Graph Labeled C

of the resources they describe. Figure 4's diagrammatic simplicity becomes problematic when large quantities of resources are to be described, when the number and kinds of relationships recorded grows large, and when more comprehensive but less-detailed views of bibliographic relationships are desired. To address these problems in a comprehensive fashion, we examined similar complex description scenarios in the sciences and borrowed another idea from the physics community—paper tool creation and use.

Paper Tools: Graph-Aware Diagram Creation

Paper tools are collections of symbolic elements (diagrams, characters, etc.), whose construction and manipulation are subject to specified rules and constraints.²³ Berzelian chemical notation (e.g., $C_6H_{12}O_6$) and—more prominently—Feynman diagrams like those in figure 5 are familiar examples of paper tool creation and use.²⁴

Creating a paper tool resource diagram requires that the rules for creating resource descriptions be reflected in diagram elements, properties of diagram elements, and drawing rules that define how diagram/symbolic elements are connected to one another (e.g., the formula $C_6H_{12}O_6$ specifies six molecules of carbon, twelve of hydrogen, and six of oxygen). The detailed bibliographic information in figure 4 is progressively schematized in a

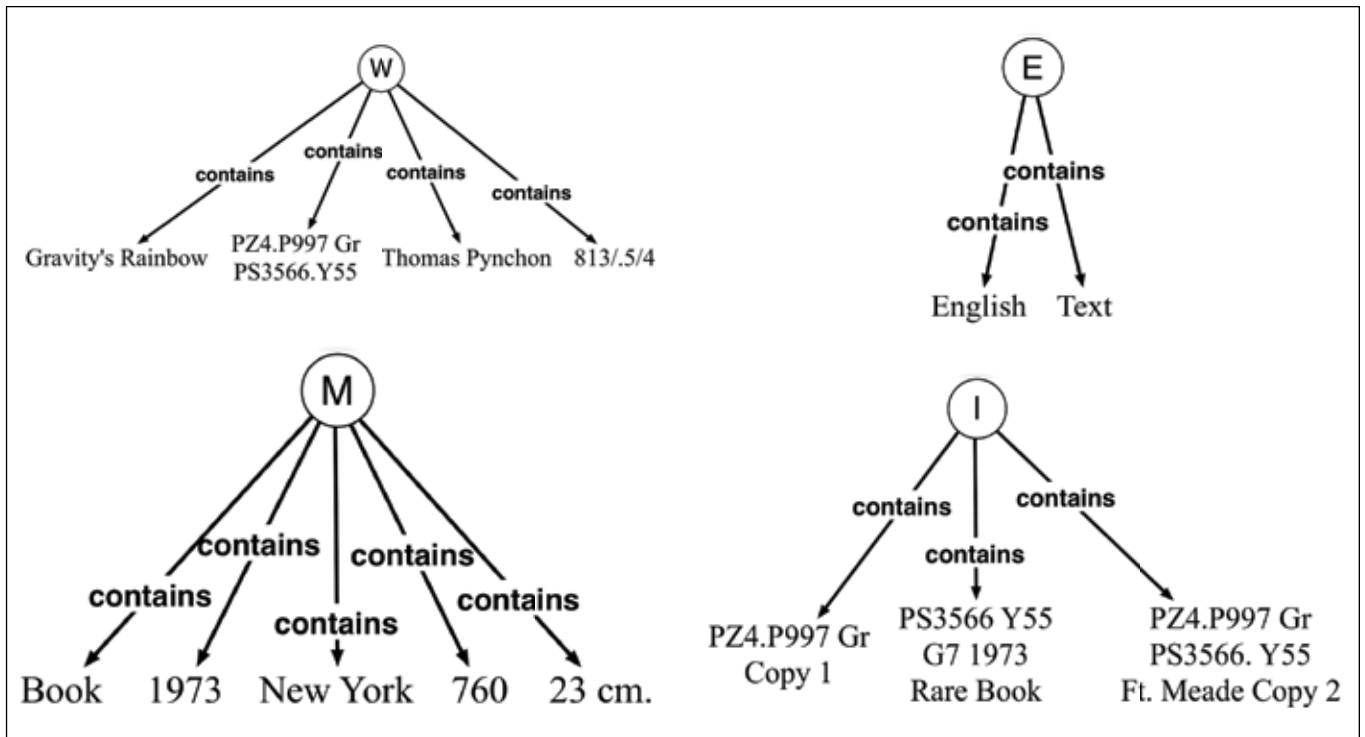


Figure 4. The All-In-One Graph in Figure 3, Separated Into Four FRBR Work (top-left), Expression (top-right), Manifestation (bottom-left), and Item (bottom-right) Graphs

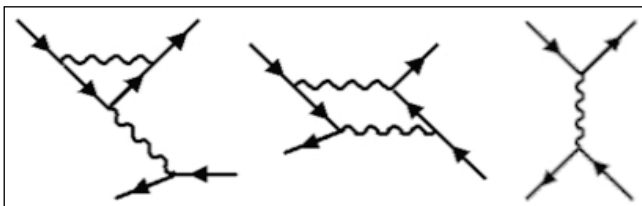


Figure 5. Feynman Diagrams of Elementary Particle Interactions

way that reflects FRBR definitions of bibliographic things of interest and their relevant relationships. As a first step, the four WEMI descriptions in figure 4 are given a common identity by linking them to a C node, as in figure 6.

The diagram is then further schematized such that FRBR description types and relationships are represented by appropriate graphical elements connected to other elements. The result shows how a FRBR paper tool makes it much easier to construct and examine complex large-scale properties of resource and resource description structures (like figure 7, right side) without being distracted by textual and linkage details.

The resource described (but not shown) by the figure

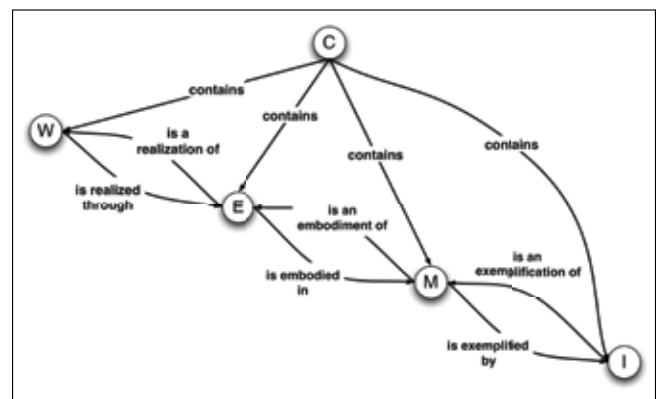


Figure 6. A FRBR Resource Description Graph

6 graph is now represented explicitly by a black dot in a ring in the more schematic paper tool version. Resource descriptions are then represented in fixed colors and positions relative to the resource/ring: the work-level resource description is represented by a blue box, Expression by a green box, manifestation by a yellow box, and item by a red box. Depicting one aspect of the FRBR

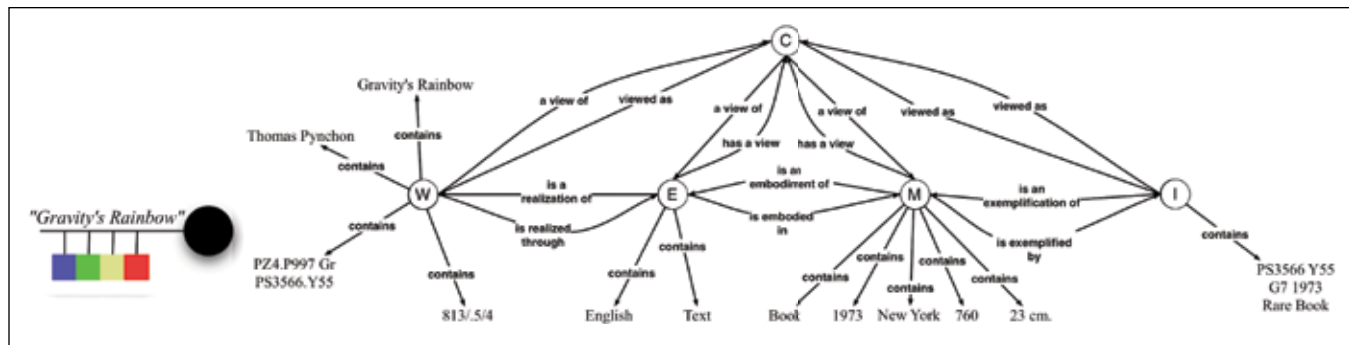


Figure 7. A FRBR Paper Tool Diagram Element (Left) and the Less Schematic FRBR Resource Description Graph It Depicts (Right)

model graphically, the descriptions closest to the black dot resource/slot are the most concrete and those furthest away the most abstract. (Readers wishing to interpret FRBR paper tool diagrams without reference to color values should note the strict ordering of WEMI elements: W–E–M–I–Resource/Ring or Resource/Ring–I–M–E–W.)

Finally, to minimize element use when pairs of WEMI boxes touch, the appropriate FRBR linking relationship for the relevant pair of descriptions (as explicitly shown in the expanded graph) is implied but not shown.

With appropriate diagramming conventions, the process of creating and exploring resource description complexes addresses combined issues of cataloging theory and institutional policy—and results in an ability to make better-informed judgments/computations about resource descriptions and their referenced resources. As a result, resource description graphs are readily created and transformed to serve theoretical—and with greater experience in thinking and programming along graph-friendly lines, practical—ends. One example of transformability would arise when exploring the implications of removing redundant portions of related resource descriptions as more copies of the same work are brought to the bibliographic universe.

The FRBR paper tool elements and the more articulated resource description graphs in figure 8 both depict the consequences of a practical act: combining resource descriptions for two copies of the same edition of the novel *Gravity's Rainbow*.²⁵ The top-most FRBR diagram and its magnified section depict how the graph would look with a single item-level description, the call number for one physical copy. The bottom-most FRBR diagram and its magnified section depict the graph with two item-level descriptions, the call numbers for two physical copies.

A FRBR paper tool's flexibility is useful for exploring potentially complex bibliographic relationships created or uncovered by scholars—parties whose expertise lies in identifying, interrelating, and discussing creative concepts and influences across a full range of communicative

expressions. The work products of scholars—especially those creations that are dense with quotations, citations, and other types of direct and derived textual and graphical reference within and beyond themselves—are excellent environments for paper tool explorations and more generally, for testing of exemplars—solutions to the potentially complex problem of describing cultural heritage resources.

Exemplars

The fourth principle in our cultural heritage resource description theory involves exemplar identification and analysis. According to the historian of science Thomas S. Kühn, exemplars are sets of concrete problems and solutions encountered during one's education, training, and work. In the sciences, exemplar-based problem finding and solving involves mastery of relevant models, builds knowledge bases, and hones problem-solving skills. Every student in a field would be expected to demonstrate mastery by learning and using their field's exemplars. Change within a scientific field is manifest by the need to modify old or create new exemplars as new problems appear and must be solved.²⁶

A cultural heritage resource description theorist would, in addition to identifying and developing exemplars from real bibliographic data and other sources, want to speculate about *possible* resource/description configurations that call for changes in existing information technologies. To the theorist, it would be as important to find out what *can't* be done with FRBR and other resource description models at library, archive, museum, and Internet scales, as it is to be able to explain routine item cataloging and tagging activities. Discovering system limitations is better done in advance by simulating uncommon or challenging circumstances than by having problems appear later in production systems.

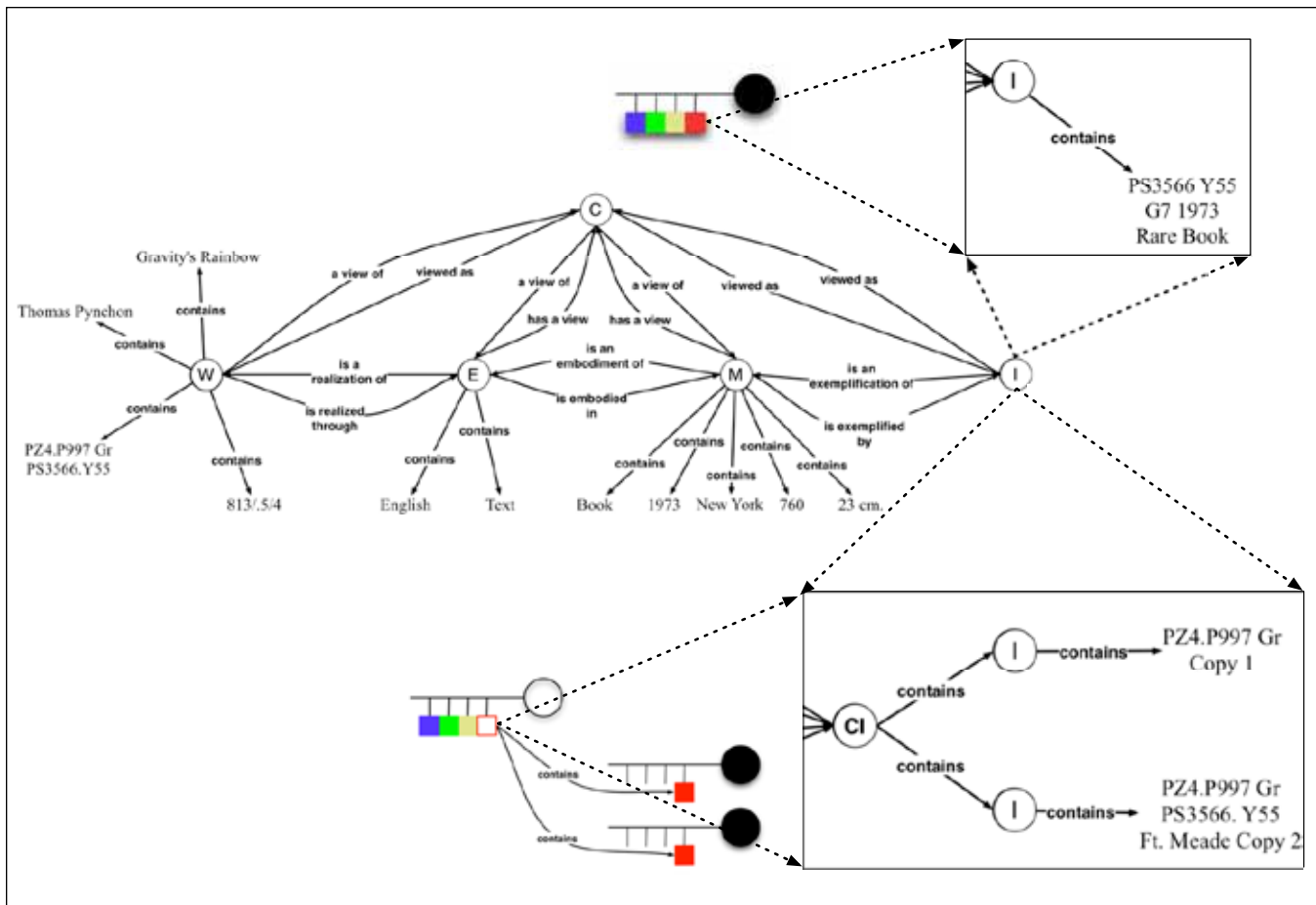


Figure 8. FRBR Paper Tool Diagram Elements and the FRBR Resource Description Graphs They Depict

Exemplars Are Not Use Cases

Use cases are a software modeling technique employed by the W3C Library Linked Data Incubator Group (LLD XG) in support of requirements specification.²⁷ Kühn-style exemplars are definitely not to be confused with use cases, which are requirements-gathering documents that contribute to software engineering projects. There is a Wikipedia definition of a use case that describes its properties:

A use case in software engineering and systems engineering, is a description of steps or actions between a user (or “actor”) and a software system which leads the user towards something useful. The user or actor might be a person or something more abstract, such as an external software system or manual process. . . . Use cases are mostly text documents, and use case modeling is primarily an act of writing text and not

drawing diagrams. Use case diagrams are secondary in use case work.²⁸

As products of and guides for theory making, resource description exemplars have different origins and audiences than those for use cases.

While use cases and exemplars offer perspectives that can support information system design, exemplars were originally introduced as theoretical entities by Kühn to explain how theories and theory-committed communities can crystallize around problem-solution sets, how these sets also can serve as pedagogical tools, and why and when problem-solution sets get displaced by new ones. The proposed process of cultural heritage exemplar creation and use, followed by modification or replacement in the face of changes in the bibliographic universe draws on Kühn’s and historian of science David Kaiser’s interest in how work gets done in the sciences, in addition to their rejection of paradigms as eerie self-directing processes.²⁹

In addition, resource description structures specified in an exemplar can and should represent a more abstract treatment of a resource description and not just data or data structures engaged by end users.

Exemplars on Hand and Others to Come

Cultural heritage resource description exemplars have been created over time as solutions to problems of resource description and later made available for use, study, mastery, and improvement. While not necessarily bound to a particular information technology, such as papyrus, parchment, index cards, database records, or RDF aggregations, resource description exemplars have historically provided descriptive solutions of physical resources whose physical and intellectual structure had originally been innovative solutions to describing, for example,

- a manuscript (individual and related multiples, published but host to history, imaginary, etc.);
- a monograph in one edition (individual and related multiples);
- a monograph in multiple editions (individual and related multiples); and
- a publication in multiple media, created sequentially or simultaneously.

With the advent of electronic and then digital communications media, more complex resource description problem-solution sets have been called for as a response to enduringly or recently more sophisticated creative/editorial decision-making and to more flexible print and digital information technology production capabilities. The most challenging problem-solution sets involve the assembly and cross-referencing of several multipart—and possibly multimedia—creative or editorially constructed works, such as the following:

- A work published as a monograph, but which has been reprinted and reedited; translated into numerous languages; supplemented by illustrations from multiple artists; excerpted and adapted as plays, an opera, comic books, and cartoon series; multimedia mash-ups; and has been directly quoted in paintings and other graphic arts productions, and has been the subject of dissertations, monographs, journal articles, etc.
- A continuing publication (individual and related multiple publications, special editions, name, publisher, editorial policy changes, etc.).
- A monograph whose main content is composed nearly entirely of excerpts from other print publications.³⁰
- A library-hosted multimedia resource and its associated resource description network.

- A webpage and its underlying, globally distributed, multimedia resource network, as it changes over time.

Such exemplars can be presented diagrammatically through the use of paper tools. This use of diagrams in support of conceptualization and information system design is deliberately patterned after professional data modeling theory and practice.³¹

Paper tool-supported analyses of a nineteenth-century American novel (exemplar 1) and of eighteenth-century French poems drawn from state archives (exemplar 2) will be presented to illustrate how information system design and pedagogy can be informed by exemplary scholarly research and publication, combined with narrativized diagrammatic representations of bibliographic and other relationships in traditional and digital media.

Exemplar 1. From *Moby-Dick* to Mash-Ups—A Print Publication History and Multimedia Mash-Up

Problem

Document the publication history of print copies of a literary work, identifying editorially driven content transfer across print editions along with content selection and transformation in support of multimedia resource creation.

Solution

The solution to this descriptive problem relies heavily on placing resource descriptions into groups and then defining relationships within and across those groups—i.e., on graph creation. After locating a checklist that documented the publication history of the novel and after identifying key components of a *Moby-Dick* and Orson Welles-themed multimedia resource appropriation and transformation network, Murray used the FRBR paper tool along with additional connection rules to create a resource description diagram (RDD) that represented G. Thomas Tanselle's documentation of the printing history (from 1851 to 1976) of Herman Melville's epic novel, *Moby-Dick*.³² The resulting diagram provides a high-level view of a large set of printed materials—depicting concepts such as a creative work, the expression of the work in a particular mode of languaging (i.e., speech, sign, image), and more concrete concepts such as publications.

To reduce displayed complexity, sets of FRBR diagram elements were collapsed into green shaded squares representing entire editions/printings, yielding figure 9.³³ The vertical axis represents the year of publication, starting with the 1851 printings at the top.

Connected Squares

The resulting network of connections in figure 9 can be interpreted in publishing terms. One line or two or more lines descending downwards from a printing's green

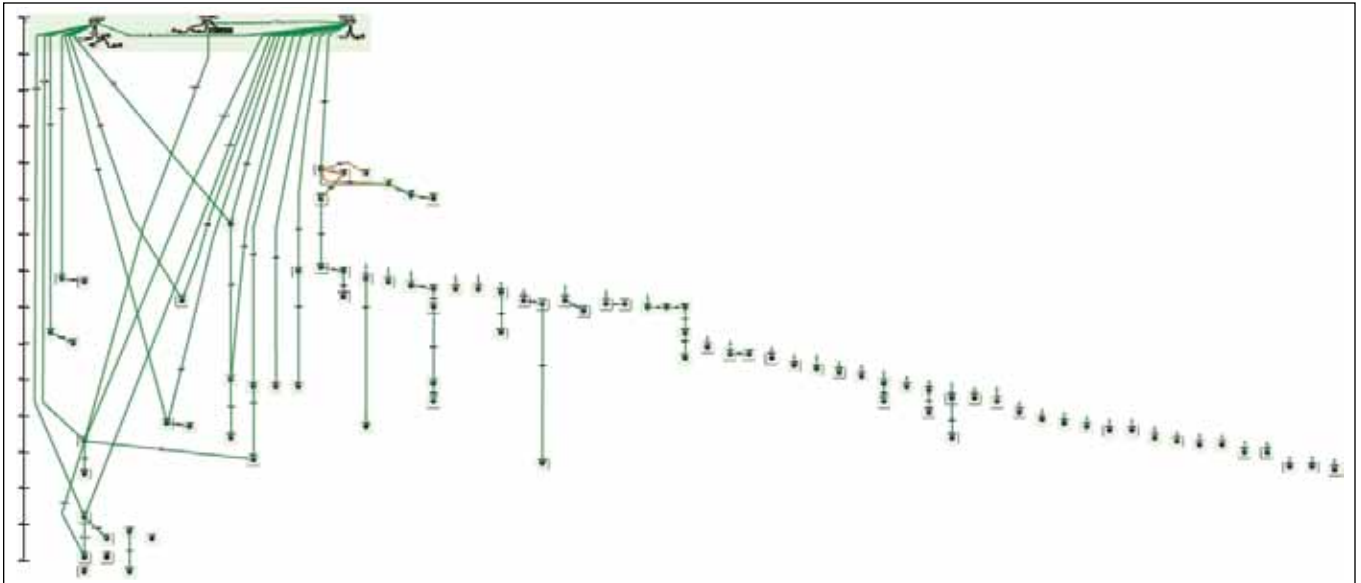


Figure 9. A *Moby-Dick* Resource Description Diagram, Depicting Relationships between Printings Made between 1851–1976 (Greatly Reduced Scale)

square are interpreted to mean that the printing gave rise to one or more additional printings, which may occur in the same or later years. Two or more lines converging on a green square from above indicate that the printing was created by combining texts from multiple prior printings—an editorial/creative technique similar to that used to construct the mash-ups published on the web.

Connecting Unconnected Squares

Tanselle’s checklist did not specify predecessor or successor relationships for each post-1851 printing. This often unavoidable, incomplete status is depicted in figure 9 as green squares that are

- not linked to any squares above it, i.e., to earlier printings; and/or
- not linked to any squares below it, i.e., to later printings; or
- connected islands, without a link to the larger structure.

Recognizing the extent of *Moby-Dick* printing disconnectedness in Tanselle’s checklist and developing a strategy for dealing with it only by analyzing Tanselle’s checklist would be extremely difficult. In contrast, the disconnectedness of the *Moby-Dick* resource description network, and its implications for search-based discovery based on following the depicted relationships is readily discernable in figure 9. The ease with which the disconnected condition can be assessed also hints at benefits to be gained by collaborative resource description supported

by paper tool diagram creation, analysis, and subsequent action, namely,

- connecting the squares (i.e., assigning at least one relationship to a printing) ensures access based on the relationship assigned; and
- parties located around the globe can examine a given connected or disconnected resource description network and develop strategies for enhancing its usefulness.

The wealth of descriptive information available in the *Moby-Dick* exemplar illustrates how previous and future collaborative efforts between cultural heritage institutions and other parties have already generated resource descriptions that possess a network structure alongside its content. With a more graph-friendly and collaborative implementation, Melville scholars, scholarly organizations,³⁴ and enthusiasts could more effectively examine, discuss, and through their actions enhance the *Moby Dick* resource description network’s documentary, scholarly, and educational value.

In its original form, the *Moby Dick* resource description diagram (and the exemplar it partially documents) only depicted full-length publications of Melville’s work. As a test of the FRBR paper tool’s ability to accommodate both traditional and modern creative expressions in individual and aggregate form—while continuing to serve theoretical, practical, and educational ends—Murray added a resource description network for *Orson Whales*,³⁵

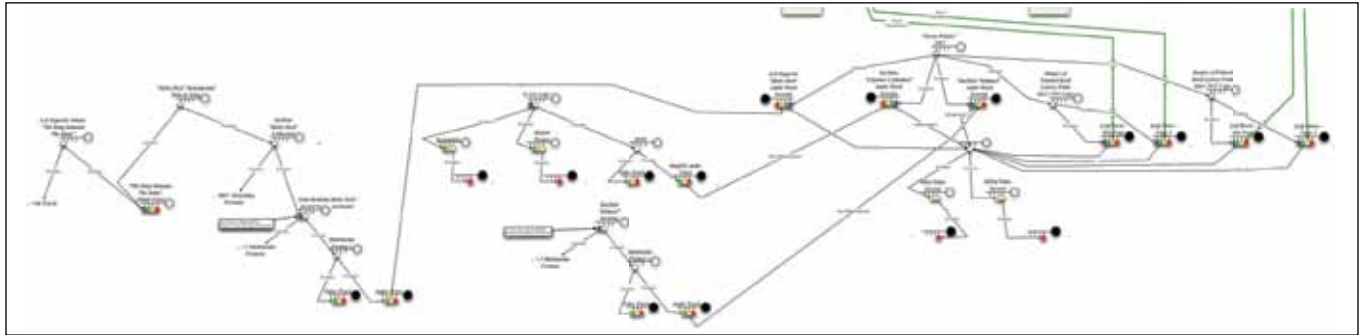


Figure 10. A Resource Description Diagram of Alex Itin's Moby-Dick Multimedia Work, Depicting the Resources and Their FRBR Descriptions.

Alex Itin's Moby-Dick-themed multimedia mash-up, to the print media diagram.

The four-minute long *Orson Whales* multimedia mash-up contains hundreds of hand-painted page images from the novel, excerpts from the Led Zeppelin song "Moby Dick," parts of two vocal performances by the actor Orson Welles, and a video clip from Welles's motion picture *Citizen Kane*. The result is shown in figure 10.³⁶

The leftmost group of descriptions in figure 10 depicts various releases of Led Zeppelin's "Moby Dick." The central group depicts the sources of two Orson Welles audio dialogues after they had been ripped (i.e., digitized from physical media) and made available online. The grouping on the right depicts the *Orson Whales* mash-up itself and collections of digital images of painted pages created from two printed copies of the novel.

Exemplar 2. Poetry and the Police—Archival Content Identification and Critical Analysis

Problem

Examine archival collections and select, describe, and document ownership and other relationships of a set of documents (poems) alleged to have circulated within a loosely defined social group.

Solution 1

In his 2010 work, *Poetry and the Police: Communication Networks in Eighteenth-Century Paris*, historian Robert Darnton studied a 1749 Paris police investigation into the transmission of poems highly critical of the French king, Louis XV. After combing state archives for police reports, finding and identifying scraps of paper once held as evidence, and collecting other archival materials, Darnton was able to construct a poetry communication network diagram,³⁷ which, along with his narrative account, identified a number of parties who owned, copied, and transmitted six of the scandalous poems and placed their activities in a political, social, and literary context.

Darnton's book can stand on its own as an exemplar for historical method, with the diagram providing additional diagrammatic support.

Solution 2

Darnton's analysis treated each poem found in the archives as an individual creative work,³⁸ enabling the use of the FRBR paper tool (as a bookkeeping device this time) instead of a tool designed to aggregate and describe archival materials. The resulting diagram is a more articulated FRBR paper tool depiction of Darnton's poetry communication network, a section of which appears as figure 11.

The depiction of the poetry communication network shown in figure 11 is composed of:

- Tan squares that depict *individuals* (clerks, professors, priests, students, etc.) who read, discussed, copied, and passed along the poems.
- Diagram elements that depict poetry written on scraps of paper (treated as *resources*) that were police custody, were admitted to having existed by suspects, or assumed to have existed by the police. If one's theory and business rules permit it, paper tool drawing conventions can depict descriptions of lost and nonexistent but nonetheless describable resources.
- Arrowed lines that represent *relationships* between a poem and the individuals who owned copies, those who created or received copies of the poem, etc.³⁹

With Darnton's monograph to provide background information regarding the historical personages involved, relationships between the works and the people, document selection from archival *fonds*, and the point of view of the scholar, the resulting problem-solution set can:

- Serve as enhanced documentation for Darnton-style communication network analysis and discussion.
- Serve as an exemplar for catalogers, scholars, and

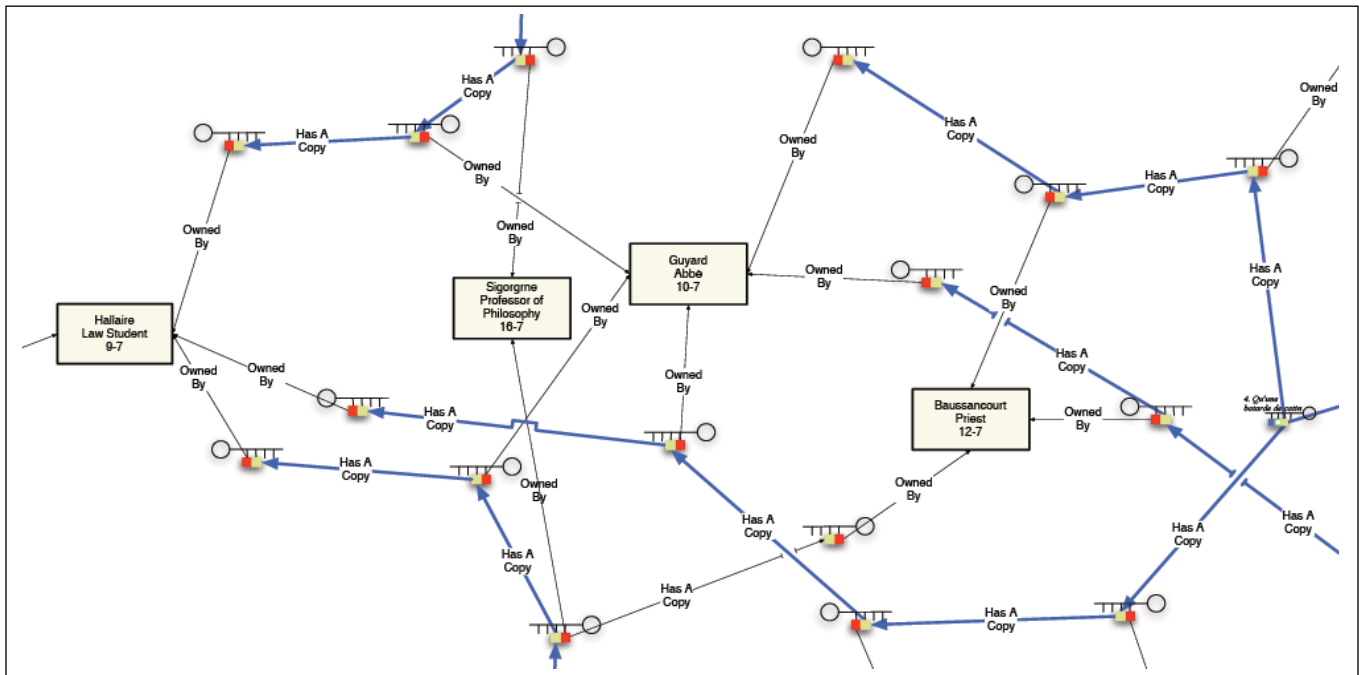


Figure 11. A Section of Darnton's Poetry Communication Network

others who seek similar solutions to their problems with identifying, describing, depicting, and discussing as individual works documents ordinarily bundled within hierarchically structured archival fonds at multiple locations.

A Paper Tool into a Power Tool

There are limits to what can be done with a hand-drawn FRBR paper tool. While Murray was able to depict large-scale bibliographic relationships that probably had not been observed before, he was forced to stop work on the *Moby-Dick* diagram because much of the useful information available could not fit into a static, hand-drawn diagram. We think that automated assistance in creating resource description diagrams from bibliographic records is required. With that capability available, cataloging theorists and parties with scholarly and pedagogical interests could interactively and efficiently explore how scholars and sophisticated readers describe significant quantities of analog and digital resources.

It would then be possible and extremely useful to be able to initiate a scholarly discussion or begin a lecture by saying, "Given a *Moby-Dick* resource description network..." and then proceed to argue or teach from a diagram depicting all known printings of *Moby-Dick*—along

with all of the adaptations and excerpts extant within a specified bibliographic universe (such as the cataloging records that appear in OCLC's WorldCat bibliographic database). Resource description diagrams, created from real-world or theoretically motivated considerations, would then provide a diagrammatic means for depicting the precise and flexible underlying mathematical ideas that, heretofore unrecognized but nonetheless systematically employed, serve resource description ends. If the structure of a well-motivated and constructed resource description diagram subsequently makes data representation and management requirements that a given information system cannot accommodate, cataloging theorists and information technologists alike will then know of that system's limitations, will work together on mitigating them, and will embark on improving system capabilities.

Cataloging Theory, Tool-Making, Education, and Practice

This modernized resource description theory offers new and enhanced roles and benefits for cultural heritage personnel as well as for the scholars, students, and those members of the general public who require support not just for searching, but also for collecting, reading, writing, collaborating, monitoring, etc.⁴⁰ Information systems that

couple modern, high-level understandings about how cultural heritage resources can be described, organized, and explored with data models that support linking within and across multiple points of view will be able to support those requirements.

The Complementarity of Cosmological and Quantum-Level Views

Cataloging theory formation and practice—two areas of activity that did not interest many outside of cultural heritage institutions—can now be understood as a much more comprehensive multilayered activity that is approachable from at least two distinct points of view. The approach presented in this paper represents a *cosmological-level view* on the bibliographic universe. This treatment of existing or imaginable large-scale configurations of cultural heritage resource descriptions serves as a complement to the *quantum-level view* of resource description, as characterized by IT-related specificities such as character sets, identifiers, RDF triples, triplestores, etc.

Activities at the quantum level—the domain of Semantic Web technologists and others—yield powerful and relatively unconstrained information management systems. In the absence of cosmological-level inspiration or guidance, these systems have not necessarily been tested against nontrivial, challenging cultural heritage resource description scenarios like those documented in the above two exemplars. Applying both views to the bibliographic universe would clearly be beneficial for all institutional and individual parties involved. If ever a model for multilevel, multidisciplinary effort was required, the history of physics is illuminated by mutually influential interactions of cosmological- and quantum-level theories, practices, and pedagogy. Workers in cultural heritage institutions and technologists pursuing W3C initiatives would do well to reflect on the result.

Ready for the Future—and Creating the Future

To explore the cultural, scientific, and mathematical ideas underlying cultural heritage resource description, to identify, study, and teach with exemplars, and to exploit the theoretical reach and bookkeeping capability of paper tool-like techniques is to pay homage to the cultural heritage community's 170+ year-old talent for pragmatic, implementation-oriented thinking, while at the same time pointing out a rich set of possibilities for enhanced service to society. The cultural heritage community can draw inspiration from geometrician Bernhard Riemann's own justification for his version of thinking outside of the box called Euclidean geometry:

The value of non-Euclidean geometry lies in its ability to liberate us from preconceived ideas in preparation for the time when exploration of physical laws might demand some geometry other than the Euclidean.⁴¹

Taking Riemann to heart, we assert that the value of describing cultural heritage resources as observations organized into graphs and of enhancing and supplementing the resource description exemplars that have evolved over time and circumstance rests in opportunities for liberating the cultural heritage community from preconceived ideas about resource description structures and from longstanding points of view on those resources. Having achieved such a goal, the cultural heritage community would then be ready when the demand came for resource description structures that must be more flexible and powerful than the traditional ones. Given the unprecedented development of the web and the promise of bottom-up Semantic Web initiatives, we think that the time for the cultural heritage community's liberation is at hand.

Acknowledgments

The authors wish to thank Beacher Wiggins and Dianne van der Reyden, directors of the Library of Congress Acquisitions and Bibliographic Access Directorate and the Preservation Directorates, respectively, for supporting the authors' efforts to explore and renew the scientific and mathematical foundations of cultural heritage resource description. Thanks also to Marcia Ascher, David Hay, Robert Darnton, Daniel Huson, and Mark Ragan, whose scholarship informed our own; and to JoAnne O'Brien-Levin for her critical eye and for editorial advice.

References and Notes

1. OED Online, "catalogue, n." <http://www.oed.com/viewdictionaryentry/Entry/28711> (accessed Aug. 10, 2011).
2. Peter Galison, "Part II: Building Data," in *Image & Logic: A Material Culture of Microphysics* (Chicago: Univ. of Chicago Pr., 2003): 370–431.
3. Gordon McQuat, "Cataloguing Power: Delineating 'Competent Naturalists' and the Meaning of Species in the British Museum," *British Journal for the History of Science* 34, no. 1 (Mar. 2001): 1–28. Exclusive control of classification schemes and of the records that named and described its specimens are said to have contributed to the success of the British Museum's institutional mission in the nineteenth century. As a division of the British Museum, the British Library appears to have incorporated classification concepts (hierarchical structuring) from its parent and elaborated on the Museum's strategies for cataloging species.
4. OED Online, "observation, n." <http://www.oed.com/viewdictionaryentry/Entry/129883> (accessed July 8, 2011).

5. David C. Hay, *UML and Data Modeling: A Vade Mecum for Modern Times* (Bradley Beach, N.J.: Technics Pr., forthcoming 2011): 124–25. Some scholars argue that decisions as to what the things of interest are and the categories they belong to are influenced by social and political factors. Geoffrey C. Bowker, Susan Leigh Star, *Sorting Things Out: Classification and Its Consequences* (Cambridge, Mass.: MIT Pr., 1999).

6. Gerald Holton, "The Roots of Complementarity," *Daedalus* 117, no. 3 (1988): 151–97, <http://www.jstor.org/stable/20023980> (accessed Feb. 24, 2011).

7. Niels Bohr, quoted in Aage Petersen, "The Philosophy of Niels Bohr," *Bulletin of the Atomic Scientists* 19, no. 7 (Sept. 1963): 12.

8. Niels Bohr, "Quantum Physics and Philosophy: Causality and Complementarity," in *Essays 1958–1962 on Atomic Physics and Human Knowledge* (Woodbridge, Conn.: Ox Bow, 1997): 7.

9. For cataloging theorists, the description of cultural heritage things of interest yields groups of statements that occupy different levels of abstraction. Upon regarding a certain physical object, a marketer describes product features, a linguist enumerates utterances, a scholar perceives a work with known or inferred relationships to other works, and so on.

10. Marcia Ascher, *Ethnomathematics: A Multicultural View of Mathematical Ideas* (Pacific Grove, Calif.: Brooks/Cole, 1991); Ascher, *Mathematics Elsewhere: An Exploration of Ideas across Cultures* (Princeton: Princeton Univ. Pr., 2002).

11. A timeline of events, people, and so on that have had or should have had an impact on describing cultural heritage resources is available online. Seven fields or subfields are represented in the timeline and keyed by color: Library & Information Science; Mathematics; Ethnomathematics; Physical Sciences; Biological Sciences; Computer Science; and Arts & Literature. Ronald J. Murray, "The Library Organization Problem," *Dipity.com*, Aug. 2011, <http://www.dipity.com/rmur/Library-Organization-Problem/> or <http://www.dipity.com/rmur/Library-Organization-Problem/?mode=fs> (fullscreen view).

12. Barbara Ann Barnett Tillett, "Bibliographic Relationships: Toward a Conceptual Structure of Bibliographic Information Used in Cataloging" (PhD diss., University of California, Los Angeles, 1987); Elaine Svenonius, *The Intellectual Foundation of Information Organization* (Cambridge, Mass.: MIT Pr., 2000): 32–51. Svenonius's definition is opposed to database implementations that permitted Boolean operations on records at retrieval time.

13. Ronald J. Murray, "The Graph-Theoretical Library," *Slideshare.net*, July 5 2011, <http://www.slideshare.net/RonMurray/-the-graph-theoretical-library>.

14. Francis J. Witty, "The Pinakes of Callimachus," *Library Quarterly* 28, no. 1–4 (1958): 132–36.

15. Ronald J. Murray, "Re-Imagining the Bibliographic Universe: FRBR, Physics, and the World Wide Web," *Slideshare.net*, Oct. 22 2010, <http://www.slideshare.net/RonMurray/frbr-physics-and-the-world-wide-web-revised>.

16. For an overview of the technology-driven Library Linked Data initiative, see <http://linkeddata.org/faq>. Murray's analyses of cultural heritage resource descriptions may be explored in a series of slideshows at <http://www.slideshare.net/RonMurray/>.

17. Pat Riva, Martin Doerr, and Maja Žumer, "FRBROO: Enabling a Common View of Information from Memory Institutions," *International Cataloging & Bibliographic Control* 38, no. 2 (June 2009): 30–34.

18. The prospects for creating graph-theoretical functions that operate on resource description networks are extremely promising. For example, *Combinatorica* (an implementation of graph theory concepts created for the computer mathematics application *Mathematica*) is composed of more than 450 functions. Were cultural heritage resource description networks to be defined using this application's graph-friendly data format, significant quantities of *Combinatorica* functions would be available for theoretical and applied uses; Siriam Pemmaraju and Steven Skiena, *Computational Discrete Mathematics: Combinatorics and Graph Theory with Mathematica* (New York: Cambridge Univ. Pr., 2003).

19. Dénes König, *Theory of Finite and Infinite Graphs*, trans. Richard McCoart (Boston: Birkhäuser, 1990); Fred Buckley and Marty Lewinter, *A Friendly Introduction to Graph Theory* (Upper Saddle River, N.J.: Pearson, 2003); Oystein Ore and Robin Wilson, *Graphs and Their Uses* (Washington D.C.: Mathematical Association of America, 1990).

20. Leonhard Euler, "Solutio Problematis Ad Geometriam Situs Pertinentis," *Commentarii Academiae Scientiarum Imperialis Petropolitanae* no. 8 (1736): 128–40.

21. "Set Theory, branch of mathematics that deals with the properties of well-defined collections of objects, which may or may not be of a mathematical nature, such as numbers or functions. The theory is less valuable in direct application to ordinary experience than as a basis for precise and adaptable terminology for the definition of complex and sophisticated mathematical concepts." Quoted from *Encyclopædia Britannica Online*, "Set Theory," Oct. 2010, <http://www.britannica.com/EBchecked/topic/536159/set-theory> (accessed Oct. 27, 2010).

22. IFLA Study Group on the Functional Requirements for Bibliographic Records, *Functional Requirements for Bibliographic Records: Final Report* (Munich: K.G. Saur, 1998). This document is downloadable as a PDF from <http://www.ifla.org/VII/s13/frbr/frbr.pdf> or as an html page at <http://www.ifla.org/VII/s13/frbr/frbr.htm>.

23. Ursula Klein, ed., *Experiments, Models, Paper Tools: Cultures of Organic Chemistry in the Nineteenth Century* (Stanford, Calif.: Stanford Univ. Pr., 2003); Klein, ed., *Tools and Modes of Representation in the Laboratory Sciences* (Boston: Kluwer, 2001); David Kaiser, *Drawing Theories Apart: The Dispersion of Feynman Diagrams in Postwar Physics* (Chicago: Univ. of Chicago Pr., 2005).

24. For more examples and a general description of Feynman diagrams, see <http://www2.slac.stanford.edu/vvc/theory/feynman.html>.

25. An enlarged version of this diagram may be found online. Ronald J. Murray and Barbara B. Tillett, "FRBR Paper Tool Diagram Elements and the FRBR Resource Description Graphs They Depict," Aug. 2011, <http://arizona.openrepository.com/arizona/bitstream/10150/139769/2/Fig%208%20FRBR%20Paper%20Tool%20Elements%20And%20Graphs.pdf>. Other informative illustrations also are available. Murray and Tillett, "Resource Description Diagram Supplement to 'Cataloging Theory in Search of Graph Theory and Other Ivory Towers. Object: Cultural Heritage Resource Description Networks,'" Aug. 2011, <http://hdl.handle.net/10150/139769>.

26. Thomas S. Kühn, *The Structure of Scientific Revolutions*, 2nd ed. (Chicago: Univ. of Chicago Pr., 1970).

27. Daniel Vila Suero, "Use Case Report," World Wide Web Consortium, June 27, 2011, <http://www.w3.org/2005/Incubator/ldd/wiki/UseCaseReport>.

28. Wikipedia.org, "Use Case," June 13, 2011, http://en.wikipedia.org/wiki/Use_case.
29. Kaiser, *Drawing Theories*, 385–86.
30. Prime examples being Jacques Derrida's typographically complex 1974 work *Glas* (Univ. of Nebraska Pr.), and *Reality Hunger: A Manifesto* (Vintage), David Shield's 2011 textual mash-up on the topic of originality, authenticity, and mash-ups in general.
31. Graeme Simsion, *Data Modeling: Theory and Practice* (Bradley Beach, N.J.: Technics, 2007): 333.
32. Herman Melville, *Moby-Dick* (New York: Harper & Brothers; London: Richard Bentley, 1851). *Moby-Dick* edition publication history excerpted from G. Thomas Tanselle, *Checklist of Editions of Moby-Dick 1851–1976. Issued on the Occasion of an Exhibition at the Newberry Library Commemorating the 125th Anniversary of its Original Publication* (Evanston, Ill.: Northwestern Univ. Pr.; Chicago: Newberry Library, 1976).
33. Ronald J. Murray, "From Moby-Dick to Mash-Ups: Thinking about Bibliographic Networks," Slideshare.net, Apr. 2011, <http://www.slideshare.net/RonMurray/from-moby-dick-to-mashups-revised>. The Moby-Dick resource description diagram was presented to the American Library Association Committee on Cataloging: Description and Access at the ALA Annual Conference, Washington D.C., July 2010.
34. The Life and Works of Herman Melville, Melville.org, July 25, 2000, <http://melville.org>.
35. The New York artist Alex Itin describes his creation: "It is more or less a birthday gift to myself. I've been drawing it on every page of Moby Dick (using two books to get both sides of each page) for months. The soundtrack is built from searching 'moby dick' on YouTube (I was looking for Orson's Preacher from the the [sic] John Huston film) . . . you find tons of Led Zep [sic] and drummers doing Bonzo and a little Orson . . . makes for a nice Melville in the end. Cinco [sic] de Mayo I turn Forty. Ahhhhhh the French Champagne." Quoted from Alex Itin, "Orson Whales," YouTube, Jan. 2011, http://www.youtube.com/watch?v=2_3-gEm6O_g.
36. The multimedia mash-up in figure 10 was linked to the much larger Moby-Dick structure depicted in figure 9. The combination of the two yields figure 10a, which is too detailed for printout but which can be downloaded for inspection as the following PDF file: Ronald J. Murray and Barbara B. Tillett, "Transfer and Transformation of Content Across Cultural Heritage Resources: A Moby-Dick Resource Description Network Covering Full-Length Printings From 1851–1976*," July 2011, <http://arizona.openrepository.com/arizona/bitstream/10150/136270/4/Fig%2010a%20Orson%20Whales%20In%20Moby%20Dick%20Context.pdf>. In the figure, two print publications have been expanded to reveal their own similar mash-up structure.
37. Robert Darnton, *Poetry and the Police: Communication Networks in Eighteenth-Century Paris* (Cambridge, Mass.: Belknap Pr. of Harvard Univ. Pr., 2010): 16.
38. Ronald J. Murray in a discussion with Robert Darnton, Sept. 20, 2010. Darnton considered the poems retrieved from the archives as distinct intellectual creations, which permitted the use of FRBR diagram elements for the analysis. Otherwise, a paper tool with diagram elements based on the archival descriptive standard ISAD(G) would have been used. Committee on Descriptive Standards, *ISAD (G): General International Standard Archival Description* (Stockholm, Sweden, 1999–).
39. The complete poetry communication diagram may be viewed at <http://arizona.openrepository.com/arizona/bitstream/10150/136270/6/Fig%2011%20Poetry%20Communication%20Network.pdf>.
40. Carole L. Palmer, Lauren C. Teffeau, and Carrie M. Pittman, *Scholarly Information Practices for the Online Environment: Themes from the Literature and Implications for Library Science Development* (Dublin, Ohio: OCLC Research, 2009), <http://www.oclc.org/programs/publications/reports/2009-02.pdf> (accessed July 15, 2011).
41. G. F. B. Riemann, quoted in Marvin J. Greenberg, *Euclidean and Non-Euclidean Geometry: Development and History* (New York: Freeman, 2008): 371.