Chapter 3

Genres without Writers: Information Systems and Distributed Authorship

Melanie Feinberg

Abstract

Purpose – This essay demonstrates how information systems — collections of documents, data, or other information-bearing objects — function internally as sites for creative manipulation of genre resources. In the information systems context, these textual activities are not clearly traced to the purposeful actions of specific writers.

Findings – Genre development for information systems can result from actions that may appear individually to be rote, repetitive, passive, and uninteresting. But as these actions are aggregated at increasing scales, genre components interact and shift, even if change is limited to one element of the larger assemblage. Although these changes may not be initiated by writers in accordance with targeted work activities and associated rhetorical goals, the composite texts thus produced are nonetheless powerful documents that come to partially constitute the broader activities they appear to merely support.

Originality/value – In demonstrating "writerless" phenomena of genre change in distributed, regulated systems, this essay complements and extends the strong body of existing work in genre studies that emphasizes the writer's perspective and agency in its accounts of genre development. By showing how continually evolving compound documents such as digital libraries constitute such sites of unacknowledged

Genre Theory in Information Studies
Studies in Information, Volume 11, 43–66

ISSN: 2055-5377/doi:10.1108/S2055-537720140000011004

genre change, this essay demonstrates how the social actions that these composite documents facilitate for their users also change.

Keywords: Genre; information studies; metadata; cataloging; digital libraries; standards

In the 30 years since Carolyn Miller's landmark essay reanimated the notion of genre as a form of social action, the focus of genre studies research has shifted from the text in isolation to the context in which a text is produced (Miller, 1984). Current genre theory emphasizes the social processes by which genres develop, stabilize, and come to partially constitute activities. Historical studies, including Bazerman (1988) and Yates (1989), show how textual forms such as the academic journal article (Bazerman) and corporate memo and report (Yates) arise in response to particular social situations, and then, over time, come themselves to shape and regulate their associated enterprises. Scientific research experiments are designed to accommodate the accepted structural elements of journal articles, for example, because adherence to established genre conventions (such as clearly defined article sections for data collection methods) facilitates acceptance in peer review. Similarly, business outcomes are quantified according to metrics that are easy to express via the quickly scanned charts and graphs of executive summaries in business reports, and outcomes that are not amenable to such measurement accordingly become difficult to justify. In this approach to genre, writing knowledge involves more than the ability to construct precise, grammatically correct sentences and paragraphs. Expert writers are in a sense socially skilled; they are able to appropriately manipulate genre conventions to accomplish their rhetorical goals (Bawarshi, 2003). Genre mastery is ultimately demonstrated through the ability of texts to serve particular functions and facilitate the activities in which they are embedded.

Extending this socially focused understanding of genre, one area of research has explored the ways in which multiple genres intersect to support and structure complex activities. Concepts such as genre sets, repertoires, systems, and ecologies have been advanced to consider how genres are deployed in both sequential and overlapping ways (Spinuzzi, 2004). To take full account of the activities structured by such genre assemblages, researchers have paid greater attention to informal, personal, and ephemeral forms of documents, such as to-do lists and scribbled annotations, and have placed less emphasis on traditional forms of writing, such as journal articles and business reports. To compare the various frameworks associated with genre assemblages (sets, systems, repertoires, ecologies),

Spinuzzi presents the example of Ralph, a fictional employee at a telecommunications company. Ralph uses calendars, notebooks, sticky notes, customer lists, account databases, faxes, and more to accomplish the simple task of securing a customer's late payment. Although Ralph uses a vast array of document genres in his work activities, his employment of these genres does not involve the kinds of formal writing skills associated with academic articles and corporate reports. Ralph writes names and phone numbers on a list, and annotates customer records with notes. Many of these textual acts, Spinuzzi suggests, are not primarily communicative in nature (they aren't directed toward an audience) but are rather mediating (they support tasks).

If recent genre research has expanded to encompass many varieties and purposes of *writing*, however, it has remained oriented around *writers* as the purposeful initiators of genre development — even if these writers don't describe the work they do as writing, nor do they present their innovations as skilled accomplishment. Spinuzzi's scenario with Ralph, for example, focuses around the various texts he uses and produces to accomplish his work tasks. Although Ralph refers to customer records in databases, Spinuzzi frames these documents through Ralph's perspective as the primary actor, or writer: Ralph adds notes to a record, for example, changing its content. Ralph has agency to shape the genres he uses, potentially innovating new genres or initiating changes in existing ones.

In contrast, the database in Spinuzzi's example functions much like a pad of sticky notes: both are merely environments to support Ralph's writing, which in turn supports his work activities. In a subsequent study of analysts at Semoptco, a search engine optimization (SEO) firm, Spinuzzi notes the significant role of a newly developed in-house information system in the preparation of client reports. Spinuzzi's interest in the information system lies primarily in the work that analysts perform to create new types of charts and graphs that the system doesn't currently support (Spinuzzi, 2010). Spinuzzi's analysis implies that the textual work involved in producing a new chart is creative and interesting: it constitutes the rapid development of new genres in response to constantly changing business conditions. For Spinuzzi, once a textual action, such as the instantiation of a particular kind of chart, becomes automated through the firm's information system, it has become regularized and stable, and, from his perspective, not interesting as the object of study. Spinuzzi keeps his attention focused on the analysts (writers) and the textual products (reports) that result from their continuous, creative innovation of genre resources. When Spinuzzi looks at the SEO firm's information system, which he calls BRILLIANCE, he sees a kind of document factory that produces information components for the writers (analysts) to creatively manipulate. BRILLIANCE takes over the rote, repetitive aspects of the writers' work. BRILLIANCE itself is not

understood as a site of dynamic, ongoing innovation and flexible deployment of genre resources.

In this chapter, I demonstrate how information systems — collections of documents, data, or other information-bearing objects, including Spinuzzi's example of BRILLIANCE — function internally as sites for creative manipulation of genre resources. However, in the information systems context, these textual activities are not clearly traced to the purposeful actions of specific writers. Genre development for information systems can result from actions that may appear individually to be rote, repetitive, passive, and uninteresting. But as these actions are aggregated at increasing scales, genre components interact and shift, even if change is limited to one element of the larger assemblage. Although these changes may not be initiated by writers in accordance with targeted work activities and associated rhetorical goals, the composite texts thus produced are nonetheless powerful documents that come to partially constitute the broader activities they appear to merely support.

Genre studies have long-considered cumulative effects of what might initially appear to be small changes over time; foundational authors such as Miller, Bazerman, and Yates emphasize the barely perceptible transitions of genre as social situations change gradually over long periods. Minor deviations enacted by individual authors are retained and aggregated by others in a dispersed process, as a tumbleweed gathers constituent plant matter into itself over its journey, losing some elements of its mass as it travels and gaining others. The more typical emphasis in genre studies, however, has been to trace and examine those changes as evidence of writer innovation within the limited flexibility of genre-mediated social action. This approach facilitates the education of skilled writers, a key concern of genre studies, which has been associated with the academic disciplines of composition and rhetoric and functional linguistics (Bawarshi, 2003; Bhatia, 2004). Accordingly, genre studies has been less likely to investigate document features that as they lie apparently dormant in realm of stabilization: elements that seem defined by regulation, of which automated systems like BRILLIANCE are an example.

In a 2011 essay that reconsiders the Semoptco case via Manuel Castells's distinction between generic and self-programmable labor, Spinuzzi clarifies this emphasis (Spinuzzi, 2011). Generic labor, which in the Semoptco case includes activities such as bookmarking Web sites with social media tools and generating automated "report cards" of search engine optimization keyword performance, is routine, low-skilled, and requires little operator discretion. From a standard genre perspective, activities that can be accomplished through generic labor can involve either official, or highly regulated, genres, or they can involve unofficial, more flexible genres. In contrast, self-programmable labor, which in the Semoptco example

includes activities such as developing a new form of competitor table and creating client action items, is autonomous, less formalized, and requires operator judgment. Activities that require self-programmable labor can also involve both formal and informal genres. In Spinuzzi's analysis, the emphasis is on the potential for automation of tasks that can be fulfilled through generic labor to increase the opportunity for innovative developments for those activities that require self-programmable labor. Spinuzzi suggests that when tasks evolve to be accommodated with generic labor and official genres, as with the Semoptco "report cards" that are generated through database queries, more effort can be spent in skilled, creative tasks, such as the creation of client action items (which may involve an official, regularized genre, but is nonetheless a skilled and creative task that requires writer judgment). In other words, when tasks become amenable to generic labor and official genres, they become opportunities to be automated and subsequently "black boxed," forgotten in themselves. While these "black boxes" can form the basis for purposeful innovation of new document forms, such as new interfaces for database content, the "black box" itself does not seem interesting to open. As a complementary example, Spinuzzi refers to an example in which the "black boxing" of accident data entry and basic queries as the Iowa Department of Transportation moved from a paper system to a database enabled the development of more sophisticated data query tools and features. The development of new tools is an interesting genre situation, but the continued aggregation of data in the database remains in the black box. From the perspective of rhetoric and composition studies, this orientation on writer agency is not just appropriate and sensible, but socially responsible: by automating tasks that require generic labor, people are free to exploit their talents and expertise in innovative ways. There is no reason to open the black box.

From an information studies perspective, however, the black box is especially worthy of attention. In this essay, I look inside the black box of automated content generation and suggest how its workings, without becoming less regulated or more "self-programmable," nonetheless form a site of genre development. These genre changes are highlighted in distributed systems such as digital libraries where the "document" that users interact with is, under the surface, the amalgamation of many component documents (for digital libraries, these components are metadata records). Genre change here results not from the actions taken to create any individual component document but from the aggregation of those components and associated changes in user access and interaction with those components. The writers of individual components are not making innovative changes in self-programmable fashion; they are following rules. However, unanticipated effects arise when various mundane differences in the creation of components are aggregated and new, unforeseen component relationships

are instantiated in the compound system. In demonstrating such phenomena of genre change in distributed, regulated systems, this essay complements and extends the strong body of existing work in genre studies that emphasizes the writer's perspective and agency in its accounts of genre development. By showing how continually evolving compound documents such as digital libraries constitute such sites of unacknowledged genre change, this essay demonstrates how the social actions that these composite documents facilitate for their users also change.

To make this argument, I first provide a more complete summary of Spinuzzi's work with the SEO analysts, demonstrating how his continued interest in writers, and how their active choices shape genre development, echoes foundational genre scholars such as Bazerman and Yates. Next, I discuss efforts within cultural heritage institutions (libraries, archives, and museums) to regulate allowable expression for their associated information systems (catalogs and collections databases) through the development and application of metadata standards. These standards, I argue, purport to "automate" content development in a manner similar to that of the SEO firm's information system, BRILLIANCE. I continue by demonstrating how, for cultural heritage information systems, actions that are undertaken as the mere application of established rules, and which might initially seem minor and unimportant, are amplified when many instances are aggregated into composite distributed systems, particularly when those instances span time, space, and institutional boundaries. I contend that these "writerless" actions, in which the apparently mundane decisions of content creators who are merely following rules are combined in distributed systems, can affect genre change just as much as the purposeful acts of writers responding to immediate exigencies. To further my argument, I show how decisions that are imagined as having no effect on the structure and content of information systems — changes to access methods across physical and digital environments — also can be seen as contributing to genre change. The shift from card-based, structured browsing for library catalogs to digital keyword search occurred without altering the structure or content of catalog records. And yet the genre of the catalog has been demonstrably affected by this apparently technical and instrumental change, without "writer" intervention.

3.1. Genre Innovation as the Product of Purposeful Writer Intervention: Spinuzzi's "Secret Sauce"

Spinuzzi's (2010) study of knowledge work in an SEO firm concentrates on the rapid pace of genre innovation that results from the SEO analysts' need

to respond to dynamic business conditions of the SEO environment. SEO firms craft strategies to ensure that their clients' Web sites are prominently featured in search results for queries with relevant keywords. As part of their services, the firm studied by Spinuzzi created monthly reports for each client to document the success of current SEO campaigns. While the enterprise of SEO has been somewhat tainted by the activities of "black hat" firms who use devious maneuvers and trickery to drive traffic to client sites (see Segal, 2011), the firm that Spinuzzi studied, "Semoptco," claimed to use sanctioned, "data-driven," "evidence-based" SEO techniques to enhance client search rankings. This "reliable" approach constituted Semoptco's "secret sauce."

To persuade clients of the continued potency of their approach, Semoptco analysts spent one-third of their work time creating client reports that convincingly established measurable outcomes for their SEO activities. Many elements of these reports were standardized in the form of templated document components, some generated through the firm's in-house system for tracking client "analytics," which Spinuzzi calls "BRILLIANCE." Some of these automatically generated elements were targeted toward specific client situations by analyst intervention, via inputting additional information into the BRILLIANCE system. Other elements, particularly for bigger, more important clients who had purchased additional services, were more extensively created and customized by the analysts.

In examining Semoptco's production of these reports, Spinuzzi focuses on these elements of dynamic customization, noting how the analysts must perform rhetorical tasks of audience analysis and ethos-building to effectively develop the most persuasive SEO strategies, corresponding metrics, and associated report elements. Although the analysts do not consider their work to involve extensive "writing," Spinuzzi makes the case that they are in fact exceedingly proficient "writers," as business conditions force them to constantly innovate new report elements. To enable this creative innovation, standard elements are quickly "templated" and automated via the BRILLIANCE system and other regulating tools.

Once a report feature becomes automated in this way, Spinuzzi suggests that it becomes regulated, or stabilized. The real "secret sauce" of Semoptco's business, according to Spinuzzi, is not automation but the flexible customization that automation enables. For Spinuzzi, the knowledge work environment of Semoptco suggests that historical studies of genre development, such as those of Bazerman and Yates, need to be updated through additional study of current, continuously changing, fast-paced workplace contexts. The evolution of academic and business genres as described by Bazerman and Yates is slow, developing over periods of years, decades, or even centuries. In contrast, Spinuzzi asserts that in

environments like Semoptco, genres and associated practices are being developed on a scale of weeks.

Although the timescale is accelerated, the story that Spinuzzi tells is quite similar to Bazerman's and Yates's historical accounts. All three studies describe how, in response to changing social environments, writers take advantage of new technologies and adapt existing textual forms to serve evolving rhetorical purposes. The protagonist of these tales remains the innovative writer, who responds to dynamic conditions by manipulating available resources in new ways. Eventually, the genre, reconfigured, evolves as the result of these persistent actions, as they become recurrent.

In the next section, I will begin to show how complicated genre systems can evolve even without specific intervention by writer-protagonists. As my example, I employ an information system, like the BRILLIANCE system described in Spinuzzi's case study. The information systems associated with cultural heritage institutions (libraries, archives, and museums) are well suited for this purpose. Cultural heritage information systems, such as library and museum catalogs, as well as archival finding aids, comprise object metadata, or systematic descriptions of institutional holdings. Cultural heritage metadata practices are both complex and well documented, with extensive professional traditions. Moreover, cultural heritage institutions, particularly libraries, have established standards and processes meant to achieve stability and consistency: although the technical implementation might differ, these standards and processes are meant to "automate" and "regulate" document production in a manner similar to that of the BRILLIANCE system. Cultural heritage metadata standards aim to facilitate the interoperability of records, so that institutions can share, exchange, and aggregate them reliably. These goals are similar to those of BRILLIANCE for the creation of interchangeable and reliably consistent report elements. In the next section, I will describe how, despite the extensive regulatory environment promoted by these various metadata standards, genre development nonetheless occurs.

3.2. Genre Regulation in Cultural Heritage Metadata: The Role of Standards

Cultural heritage institutions — libraries, archives, and museums — have a long tradition of creating documents (such as catalogs) to describe and provide access to other documents (such as the objects held by a particular museum). *Metadata* is a general term for such descriptions. Metadata facilitates searching and browsing for particular kinds of items. Metadata can comprise highly structured records, as with library catalogs, but it can also

include loosely structured narrative descriptions, such as archival finding aids. Metadata can be in paper form, as with card-based catalogs or museum inventory books, or in digital form, as with museum collections databases. Traditionally, cultural heritage metadata describes an information object and indicates how that object can be accessed (through a library call number or archival series folder, for example). Some cultural heritage information systems include both metadata and a direct access mechanism for a digital copy of the item. Information systems that provide access to both metadata and the digital objects described by the metadata are sometimes called *digital libraries*.

Andersen (2008) asserts that digital libraries and other cultural heritage information systems can be productively understood with genre theory. Andersen bases this claim on the perception that audiences, or users, of digital libraries will understand, interpret, and use them in similar ways for recurrent social purposes. Andersen explains:

> When we recognize a given digital library as such, we also know what to expect of it, that is, what it can do for users and what users can accomplish by using a digital library, as compared to other textual places on the Internet that perform similar actions. A genre understanding of digital libraries thus provides a means of understanding matters of knowledge organization, communication of symbolic activity, and information seeking. (Andersen, 2008)

Accordingly, a digital library is a genre: it is a typified response to a particular social situation. A digital library, as a form of document, enables certain actions and constrains others; its genre conventions facilitate its use for particular activities while making other potential tasks less immediately salient. A digital library is a composite genre that is formed through the aggregation of highly regulated components, the metadata records.

Quite similarly to the case of Spinuzzi's SEO analysts, the people who create the metadata that constitutes the primary content of digital libraries and other cultural heritage information systems — library catalogers, document indexers, archivists, and so on — do not typically see themselves as "writing" when they generate metadata records. Andersen argues, as Spinuzzi does, that cultural heritage information systems are nonetheless documents whose structure and content are regularized in alignment with particular social activities. One of Andersen's goals in making this case is to encourage, through awareness of digital libraries as documents, a critical stance toward their development and use (Andersen, 2006). Genre knowledge, as a form of information literacy, can help information seekers to more effectively understand and assess the nature of access enabled through cultural heritage information systems, as well as the documents provided

by them. Likewise, genre knowledge can help information professionals develop systems that support this critical attitude.

Professional practice traditions for cultural heritage metadata, however, are directly oppositional to Andersen's implication that metadata creators should approach their work as a form of writing. While Spinuzzi's SEO analysts were required by business conditions to continually develop new report elements and adapt existing ones for rapidly changing situations, professional practice for cultural heritage metadata does not endorse deviation from established conventions. Instead, excellence in metadata generation is more often located in adherence to standards. Consistency is valued over flexibility. (While consistency has been most strongly valued in libraries for the longest period of time, the digital environment, and its associated potential for aggregating cultural heritage metadata in unified systems, has encouraged increased attention to consistency for archives and museums as well.)

In Spinuzzi's case study of Semoptco, the information system BRILLIANCE produced standardized report data and other report elements. Spinuzzi does not further investigate this "automation" and how it was implemented and maintained. In cultural heritage information systems, this kind of regulation is enabled through an array of intersecting standards. These standards are meant to reduce individual variation between records and systematize metadata generation, similar to the function of BRILLIANCE. Because information objects and their attributes are quite complex, the creation of much descriptive metadata requires the skilled labor of educated catalogers; for many attributes, high-quality descriptive metadata cannot be generated algorithmically. (Computer-generated determination of subject terms for text documents, for example, remains inexact and inconsistent from item to item.) Although many aspects of metadata creation demand human judgment, however, standards are thought to guide and restrict that judgment, ensuring regulated, reliable output. Such consistency, enabled through standards, is considered necessary to achieve reliability and interoperability of records. If records are reliable, users of cultural heritage information systems will not only trust that the records are accurate; they will be able to identify, compare, and locate resources more quickly, effectively, and confidently. If records are interoperable, institutions can share them. This sharing enables work efficiencies, because a single record can serve as a template for similar items in many collections. It also increases user access, as it becomes possible to create aggregated collections with shared records from many institutions (as with the union library catalog WorldCat, which aggregates over 2 billion records from tens of thousands of libraries, or with the metadata gathered by Europeana, which aggregates 30 million records from libraries, archives, museums, and galleries across Europe). Instead of searching many library

catalogs individually to locate a specific work or set of similar works, users can search a union catalog like WorldCat once.

Gilliland (2008) categorizes the metadata standards that facilitate regulation into four basic groups. *Structure* standards define the elements (also called attributes, properties, or characteristics) by which an object is described. In a database, elements are the columns in a table. Structure standards may be called schemas. The Dublin Core schema is one example of a structure standard. Dublin Core is a simple element set meant to be appropriate for any sort of information resource. *Content* standards specify how the values associated with the structural elements are defined, expressed, and formatted. For example, Subject is a Dublin Core element (part of a structure standard) that is meant to describe what a resource is about. Content standards would specify how to determine a value for the subject (including such details as how many terms to assign, the level of specificity of those terms, their exhaustivity, and so on) as well as how to express and format the selected values. Content standards may also dictate that certain attributes should be expressed using controlled vocabularies, or restricted sets of allowable values; Gilliland calls these vocabularies *value standards*. The Library of Congress Subject Headings (LCSH) is a value standard that may be used to assign values to the Subject element in Dublin Core. (The design of these controlled vocabularies is likewise guided by associated standards for their construction; this area of standards development is not discussed by Gilliland.) Finally, *format/technical interchange* standards specify various encodings by which structure standards are implemented for computer processing. Dublin Core, for example, can be implemented as RDF/XML (the format associated with "linked data") or even as plain text. Although these categories may not always be cleanly separated in actual practice, they are useful to understand the range of standardization efforts in the cultural heritage context. Because library cataloging standards have been evolving for over a hundred years, for example, their structural basis is not as systematically defined as more recently created systems, such as Dublin Core, and specifications associated with Gilliland's structure and format categories are both encompassed within the catalog digital encoding standard, MARC (MAchine Readable Cataloging).

In the context of cultural heritage metadata, this array of standards constitutes the means by which genre conventions are regulated. Metadata standards are conceptual blueprints for "automated" content generation in the cultural heritage environment, similar to the underlying requirements by which a system like Spinuzzi's BRILLIANCE produces content elements. BRILLIANCE itself may rely more on algorithmically produced content than most cultural heritage metadata — it's unclear from Spinuzzi's description — but one can think of these metadata standards as analogous to the specifications that define algorithmic processes and

parameters. Metadata standards constrain the structure, semantics, and syntax of expression to significant degree, so that the creation of a record is perceived as a task more like scientific observation (such as identifying the material composition of a seawater sample) than creative interpretation (such as suggesting the meaning inspired by a poem's figurative language). The standards are designed to make the activity of library cataloging more like analyzing seawater and less like interpreting poetry. Accordingly, Theimer (2012) argues that emphasis on standards inhibits creativity in library cataloging:

> The average cataloging environment is not naturally conducive to creativity. It is a profession of standards and judgment based on rules that describe how data is selected and formatted. Although specific rules change, the demand to adhere to the rules has remained constant. Creating quality records, and records that may be able to migrate and crosswalk correctly, relies on consistency of data, which requires adherence to standards, both national and local. (Theimer, 2012, p. 897)

Hoffman (2009) likewise notes that while catalogers aim to be "user-centered" when creating records, they believe that following established standards is the best means to accomplish this goal, as opposed to customizing records for local circumstances. Hoffman explains that

> Catalogers also believe that cataloging standards have been crafted based on an understanding of users' needs. To catalogers, standards represent users, so to follow standards is to meet users' needs. Why do catalogers need to customize bibliographic records if the standards already meet users' needs? (Hoffman, 2009, p. 635)

In alignment with this perspective, the authorship of individual metadata records primarily involves the application of standards and associated ready-made components (such as controlled vocabularies). In determining the subject for a library item, for example, a cataloger consults the extensive rules for selecting terms from the LCSH controlled vocabulary. These rules include guidelines for subject exhaustivity and specificity (how many terms to select and their level of abstraction). The cataloger inserts these terms into a standardized place within the MARC record. The terms themselves (their spelling, capitalization, and so on) are dictated by the controlled vocabulary. While the question "what is this item about?" requires human judgment to answer, and answers will inevitably differ to some extent, the standards are meant to limit this variation as much as possible. The cataloger is not supposed to interpret the item's subject in a new, provocative, or interesting way; the cataloger is supposed to interpret

the item's subject in the most predictable, "accurate" way possible. This is precisely the sort of process enacted by one of Spinuzzi's SEO analysts when generating one of the "automated" elements supported through BRILLIANCE. The analyst makes some judgments in setting parameters and so on, but the system reduces the level of variation possible in expressing the output based on those parameters.

Despite extensive sets of standards, professional belief in the necessity of their rigorous application, and a wide range of mechanisms to ensure adherence — copious documentation and practice literature, frequent rule clarifications and updates, academic courses and professional training, and so on — the regulation of cultural heritage metadata remains imperfect. The level of variation between records created according to the same sets of standards is greater than one might anticipate given the "templated" environment. Variation is particularly apparent across time. Tennis (2012), for example, describes how the concept of "eugenics" is moved from one place to another in the Dewey Decimal Classification (DDC) across multiple editions of this controlled vocabulary, which results in books about eugenics being scattered across various numbers in the classification (it moves from biology to, in various contexts, social sciences, philosophy, and technology). The catalogers assigning the classification number for eugenics are undergoing the same process to describe books and selecting the same concept to represent the book's "aboutness"; however, the number changes over time result in these books appearing in different locations (physical and digital) in the library, with different neighbors. This kind of situation is analogous to one of Spinuzzi's analysts entering the same parameters into BRILLIANCE on multiple occasions, but the operations performed on those parameters slightly changing, as when one performance measurement is substituted for another — perhaps from number of clicks to click rate. In both situations, although each individual "writer" is performing the same regulated action — assigning a book to the current appropriate class that indicates a subject of eugenics, for the cataloging case, and demonstrating the success of SEO operations upon a particular search keyword, in the BRILLIANCE case — the expression of that act is slightly different. For generating SEO reports, however, the effects are negligible, because each report constitutes an independent communication to a specific client. For library cataloging, however, changing the underlying operation (here, the controlled vocabulary of the DDC) has a cumulative effect on the aggregated collection of records and their associated items. The books about eugenics are no longer collocated, and this affects all library users.

Although Tennis discusses eugenics because it is "a strange case," in that its many and significant movements across the DDC are unusual, Buckland (2012) asserts that the meaning of subject concepts is inherently unstable. Over and over again, terms are once used as objective descriptions are later

understood to be pejorative: examples include "Gypsies" for "Roma," and "Orientals" for "Asians." But such shifts are endemic, even when terms do not cause offense. As Buckland remarks:

> Assigned names are … inherently obsolescent with respect to both the past and the future. Discourses and the librarians flow forward with time, but the assigned names have been inscribed for, and fixed in, a receding past. (Buckland, 2012, p. 157)

When certain terms are determined to be socially inacceptable as subject descriptions, they are sometimes automatically replaced with new labels, so that all instances of "Gypsies," for example, become "Roma." But usually this does not happen, because these sorts of meaning shifts are subtle and apparently inconsequential. "Climate change" had a different meaning in the 1930s (when it likely referred to the climate changes occurring in prehistoric periods, such as ice ages, caused by natural forces) than it does now (when it more often refers to current conditions caused by human activities). But if the meaning of "climate change" as an element of descriptive metadata changes, and if this shift results in associated changes in relationships between records, causing topic scatter, these effects are "writerless"; they are caused by cascading effects of aggregation over time, and not as the result of purposeful action.

3.3. Writerless Genre Change through Aggregation: The Case of *Uncle Tom's Cabin*

The cumulative effects of such variations can be extensive. In order to constitute genre adaptation or development, however, they would need to provoke a recurrent shift in reader perception of document structure and function. To demonstrate how this might occur, I refer to the application of subject headings in the library catalog records for three manifestations of the same work: novel *Uncle Tom's Cabin*, by Harriet Beecher Stowe. Stowe, a nineteenth-century American abolitionist, wrote the novel to advocate against slavery, and the dramatic tale was extremely popular worldwide. Additionally, although Stowe opposed slavery, the black characters in the novel are now seen to embody harmful stereotypes. In particular, the term "Uncle Tom" can refer (often derogatively) to someone who passively accepts or contributes to their own oppression.

Uncle Tom's Cabin is an illustrative, but not uncommon, example of an important fictional work more known for its subject matter and historical importance than for its literary merit. Accordingly, many manifestations of *Uncle Tom's Cabin* exist, and so there are many records to compare;

additionally, subject access is demonstrably important and appropriate for *Uncle Tom's Cabin*. The subject headings for the three example editions appear in records for the library catalog at the University of Texas at Austin; these are not the only manifestations for *Uncle Tom's Cabin* in the catalog, but they represent a range of subject headings applied. (*Note:* It is not possible from the catalog itself to determine when records are made or updated, so any references to historical trends must be inferred based on publication dates of the items and cannot be firmly substantiated.)

In the following three examples, there is variation in the number and specificity of applied subject headings, with some expansion of content (as with the inclusion in Edition 3 of an access point for the character Uncle Tom). Some of these changes are likely due to the introduction of suggested (not mandatory) guidelines for providing subject access to fiction, published by the American Library Association in 1990 and revised in 2000. These guidelines supplement and extend official content standards and can be viewed, as the changes associated with eugenics in the DDC, as analogous to algorithm improvements or measurement enhancements in a system such as BRILLIANCE.

- *Edition 1 (published 1985)*.
  - Slavery — Southern States — History — Fiction.
- *Edition 2 (published 1998)*.
  - Plantation life — Fiction.
  - Southern States — Fiction.
  - African Americans — Fiction.
  - Slavery — Fiction.
  - Slaves — Fiction.
- *Edition 3 (published 2007)*.
  - Uncle Tom (Fictitious character) — Fiction.
  - Master and servant — Fiction.
  - African Americans — Fiction.
  - Fugitive slaves — Fiction.
  - Plantation life — Fiction.
  - Southern States — Fiction.
  - Slavery — Fiction.
  - Slaves — Fiction.

More interesting than the introduction of entirely new subject headings (primarily the heading for Uncle Tom as a character whose influence extends beyond his role in the novel itself) is the way that the same content is deployed differently in the three editions. The first edition's record condenses its subject description into a single faceted heading that incorporates, by implication, the ideas of all the separate headings in Edition 2,

and most of the ideas of Edition 3 (with the exception of Uncle Tom and fugitive slaves). As a concept, "Slavery – Southern States – History – Fiction" includes within itself the notions of slaves, African Americans, and, arguably, plantations (these representing, in the novel, the people enslaved and the most commonly associated site of their bondage). Editions 2 and 3 represent this coordinated, complex subject concept as a set of individual elements. Concurrently, the notion of history in Editions 2 and 3 appears only by implication when considering the individual elements as a set. ("History" would not necessarily be associated with any of the separated headings on their own, but it does arise in their combination.)

Purely from a content perspective, the changes between the metadata records for Edition 1 and Editions 2 and 3 appear minor. From Edition 1 to Edition 2, especially, we can assert inconsequential content changes as the implied aspects of Edition 1 (slaves, African Americans, and plantations) are made explicit in Edition 2, and as one explicit aspect of Edition 1 (history) is made implicit in Edition 2. From Edition 2 to Edition 3, two new content elements are added (Uncle Tom and fugitives) and another implicit content element is made explicit (masters and servants). In terms of general indexing principles, the levels of specificity and exhaustivity are not significantly modified from one edition to another. While the subsequent editions are more explicit, they are not more specific. Edition 2 does not clarify that slavery is presented from an abolitionist perspective in the novel, nor that slaves are depicted as wrongfully oppressed; Edition 3 does not indicate that the relationship between master and servant in the novel is enforced against the will of the servant. Although the structure and syntax by which the content of the subject headings varies significantly between the editions, the basic content is surprisingly stable.

However, when relationships between items are considered, the differences between the metadata records for the three editions become much greater. In the composite document of a digital library, the expression of its genre character emerges through the aggregation of individual records and accompanying relationships between the records. In the case of *Uncle Tom's Cabin*, mundane variations in the structure and syntax of subject headings for individual records results in vastly different relationships between particular editions of *Uncle Tom's Cabin* and other works in the digital library. This different network of relationships changes the character of the underlying records, the actions that can be effectively performed with them, and the user interaction with the library system. These relationship changes form the potential substance of genre development for the digital library.

In Editions 2 and 3, the subject headings must be understood in combination with each other to make sense in the context of the novel. If, for example, the heading of Plantation life – Fiction is considered on its own,

without reference to other subject headings, then it need not be associated with the antebellum South, nor with slavery. In conjunction with any of the other headings applied, however, Plantation life – Fiction takes on a particular historical context. For each metadata record, in other words, the subject headings for Editions 2 and 3 cannot be considered as independent descriptive acts but only as integrated components of a synthetic description, even as their structure implies the opposite. Conversely, Editions 2 and 3 are at the same time placed into subject relationships with items that may address the individual components of the synthetic description quite differently from *Uncle Tom's Cabin*. In the University of Texas at Austin library catalog, for example, the heading Plantation life – Fiction has been applied to novels set on tea plantations in India and sugar plantations in Haiti and Puerto Rico. Moreover, other novels in this subject group that are likewise set in the pre-Civil-War American south do not necessarily focus on slavery and its effects on the enslaved. Some of these items similarly described with Planation life – Fiction include such diverse points of view as a murder mystery focused around the artist James Audubon and a time-travel romance. Also in the University of Texas at Austin library catalog, items besides *Uncle Tom's Cabin* also described with the heading Slavery – Fiction include historical novels set in Jamaica, Cuba, and colonial Guiana, and also science-fiction novels set in the imaginary worlds of Alta and Roshar, in addition to novels set in the nineteenth-century southern United States. Significantly, the meaning of these sets is contingent, dependent upon the composition of particular collections. At the University of Texas at Austin library, Plantation life – Fiction encompasses diverse settings and perspectives on plantations. At another library, Plantation life – fiction *might* be restricted to works similar to *Uncle Tom's Cabin:* where "plantation" can *only* be the setting for the unjust and immoral slavery of African Americans. It depends on what that collection contains.

In contrast, the meaning for the composite heading of Edition 1, Slavery – Southern States – History – Fiction, is constrained. With the subject description structured this way, *Uncle Tom's Cabin* cannot be related to science-fiction novels set on Alta or mysteries featuring the artist Audubon. In the University of Texas at Austin catalog, the only items assigned to this subject heading are other editions of *Uncle Tom's Cabin*. In the WorldCat union catalog, several other novels share this heading, but they are, of course, all focused on slavery in the American south. The heading for Plantation life – Fiction, on the other hand, is assigned to over 543 novels in WorldCat, representing even more diversity amongst the included items than in the Texas catalog. For example, an extensive series of popular romance novels (Brides of Montclair) is included; these are primarily located at public libraries, not at universities.

As the structure of subject headings for fiction has changed from composite terms to sets of individual components, then, the function of the catalog also shifts, depending on the items included in any particular collection (either as part of a single institution's holdings or via aggregation across multiple institutions). In smaller, more focused collections, the subject headings provide direct access to closely aligned works. In larger, more diverse collections, the subject terms demonstrate relationships between works that may have radically different perspectives on the assigned headings. The nature of subject browsing for fiction potentially becomes quite different. When the range of items assigned to a subject heading is constrained, the headings are primarily *finding* tools: they are best for locating items that satisfy a defined information need. When the range of items assigned to a subject heading is looser, the headings become *browsing* tools, oriented toward serendipitous discovery: they facilitate the development of the information need itself.

I suggest that this type of change in supported user interaction may constitute a form of genre development, as "what users can accomplish by using a digital library," in Andersen's (2008) formulation, evolves, with browsing actions potentially replacing finding ones as results of subject-heading use. The functions enabled through the composite document of a digital library change, and with those functional changes, the genre of digital library may also change. I argue that this type of change does not result from the innovation of individual writers (in this case, catalogers) but from a change in "automation" (in this case, standards) that initially seems to be merely instrumental, or a matter of implementation details. The shift from composite headings to sets of self-contained components affects little in the context of individual catalog records. However, as these independent textual acts are aggregated across collections of various composition, the mode of user interaction enabled through the subject headings may shift, if the aggregated records are sufficiently diverse. Changes to the individual records seem — and are — relatively trivial until those changes are aggregated and the relationships between records are subsequently affected. Small, negligible changes to metadata records only become significant as they are automatically gathered together in the composite document of the digital library. In that context, however, apparently inconsequential structural technicalities lead to notable functional shifts.

3.4. Writerless Genre Change through Access Mechanism: Effects of Keyword Search

The previous section showed how the aggregation of independent acts of textual production can catalyze the evolution of a genre (here, the digital

library or catalog) without the purposeful intervention of writers. Regulated by an array of metadata standards, changes to individual records seem relatively minor even when the standards are revised over time, as the example of *Uncle Tom's Cabin* demonstrates. Evidence from the case of *Uncle Tom's Cabin* suggests that the change from a single composite subject heading to multiple discrete subject headings appears primarily syntactic in nature, similar to changing the units for a quantified measurement, as when changing the representation of density from a composite measurement to separate notation of mass and volume. However, evidence from this case study also suggests that, as these changes accumulate in collections of varying composition, the mode of access enabled through the subject headings may likewise change, from a direct finding mode to a more open browsing mode. Genre adaptation manifests with this altered function.

This section presents another site of writerless genre change, one independent of metadata record production. Here, I examine the functional evolution of the digital library (or catalog) as its primary access mechanism changes from paper card files to digital representation of the same information. Although my example case is once again drawn from library cataloging as having the longest history and greatest level of collaborative practice across institutions, similar effects may be seen also in the transition from paper to XML-based archival finding aids, and from paper to digital databases for museum collections. (The role of digitization in understanding the material element to textual interpretation has been more extensively studied in the domains of digital humanities and textual studies, where the editorial encoding of digital files became an opportunity to reexamine the material contributions to meaning. Bonnie Mak has written in this area, along with scholars such as Johanna Drucker, Jerome McGann, and Katherine Hayles, amongst many others.)

Even though a new underlying data model for catalog information, Functional Requirements for Bibliographic Records (FRBR), was endorsed by the International Federation of Library Associations (IFLA) in 1998, the structure and content of library catalog records has remained remarkably constant since the days of printed card catalogs. Although the newest revision of the cataloging content standards, Resource Description and Access (RDA), nominally endorses the newer FRBR model, RDA's continued alignment with the record structure enabled through the MARC digital format results in continued structural similarity with old paper cards, which MARC was originally created in the 1960s to duplicate (Coyle & Hillmann, 2007). In content and structure, the library catalog record as described with RDA is almost identical to the paper catalog card.

However, although in many respects the catalog record has maintained its essential integrity over the years, the mode of access for catalog

information has drastically changed as catalogs migrated to the digital environment. What might initially be described as a mere change in material instantiation of the records — from physical to digital — has had profound effects on catalog function, and these effects have proliferated as keyword searching, an access mechanism extraneous to the metadata record itself, has been introduced and increasingly emphasized in digital catalog implementations.

Conceptually, the structure of library catalog records remains focused around three primary access points of title, author, and subject. In the card catalog system, these access points were the only means by which the catalog could be searched. For each item described with a catalog record, a separate card was created for each defined access point, which was printed at the top of the card. These access points were regularized via vocabulary control so that, for example, all the works associated with a particular author were filed together under an authoritative form of name, and not (or not only) under the form of name that appeared on the item itself. Often, the cards were separated into three separate banks of filing cabinets, one for each access point type of author, title, and subject. To search the catalog, one first decided on search mode — by author, by title, or by subject — and looked up records accordingly. If one conceptualized the seeking task as looking for fictional material about plantation life, then one went to the subject file and started with the Ps, but if one conceptualized the seeking task as locating that famous book by Harriet Beecher Stowe, one went to the author file and started in with S. In this system, the structure of catalog records and the nature of the seeking task were very much integrated. In order to progress with searching, a seeker had to think about the known attributes of what was being sought and identify which of the three access modes would best support one's current conceptualization of the information need.

Current digital library catalogs still maintain this underlying structure of access points. Vocabulary control is maintained, via "authority files," for authors, subjects, and (sometimes) titles, and the basic content and structure of metadata records is remarkably similar in the current digital environment. A digital MARC record for an item is conceptually almost identical to the paper card for the "main entry," or primary record, in a card environment. Even the conceptualization of search as being defined through selected access point has been maintained. Although separate banks of cards no longer exist, of course, this function has been shifted to separate digital authority files. A pure catalog search is a two-step process: one first searches the appropriate authority file (author, title, or subject) and locates a relevant controlled vocabulary item (such as "Planatation life – Fiction" or "Stowe, Harriet Beecher, 1811–1896" or "Uncle Tom's Cabin"). One then uses that controlled vocabulary item to produce a list of

all resources associated with that access point (364 resources in the University of Texas at Austin catalog with the author "Stowe, Harriet Beecher, 1811–1896," for example). If one searches the catalog in this manner, then both the means and matter of catalog searching is exceedingly similar in the digital environment to the physical one. It is merely much faster and more convenient to locate, sort, and select between items related under an access point, and to find additional relevant access points and associated items.

Despite the continued potential for such consistent interaction modes with essentially consistent content, however, structured search via access points is no longer the predominant mode of access for library catalogs (Rose, 2012). Many users have no idea that the systematic precision enabled through the careful creation of catalog records and their rigorous specification of controlled access points even exists, a situation that was impossible with the card catalog. Instead, most users interact with the catalog via unstructured keyword search, which functions similarly to the implementation of Web search engines, such as Google. In a keyword search, all the text on a catalog record is taken as an undifferentiated bag of words. Using the keyword search for "Harriet Beecher Stowe" in the University of Texas at Austin catalog, for example, produces 632 resources (almost double that of the author search), with results that include material about Stowe as well as written by her, results included because her name is in the title, and so on.

Following a growing trend, the current primary search mechanism for the University of Texas at Austin library system does not limit its keyword matching to the library catalog itself. The relatively new "integrated discovery" system, called SCoUT, produces aggregated results across the university's catalog metadata records and across research databases for academic journal articles and other serials, so that records for individual articles are included. A keyword search for "Harriet Beecher Stowe" in SCoUT, as opposed to the library catalog proper, generates a results set of 65,439 records. The difference in scale and scope between the SCoUT results set of 65,439 items and the author search result set of 364 items is immense. Although the SCoUT interface provides a number of filters to narrow results, including a set of filters based on "subject," these filters do not provide the same window into the underlying structure of the system as either the digital or physical implementations of the catalog authority files.

Decisions regarding interface and access changes such as implementation and prioritization of keyword searching, and implementation and prioritization of integrated discovery tools like SCoUT are commonly performed independently of the cataloging process, and yet they have profound effects upon the access and interpretation of the records. Although the catalog's

structure as a precisely defined web of meticulously created relationships still exists, the refined searching and browsing functions enabled through this structure are no longer apparent to catalog users. In card form, and in its pure digital form, the catalog is a finely tuned research tool that rewards expert use with both efficient location of specific items and serendipitous discovery of aligned items. In contrast, when subsumed under broadly undifferentiated keyword search and aggregated with millions of records of individual journal articles that had previously been kept to separate research databases, the catalog becomes a lower form of Google. The catalog's pure interfaces are oriented toward producing *sets*: relatively small groups of related items that bear close scrutiny and reward analysis of their inner logic (i.e., to determine the internal relationships that indicate why these items were filed together and accordingly to select the *best* items in a set or to identify attributes for a more appropriate set). The catalog's newer, extended interfaces that result from keyword search are oriented toward producing *quick individual hits*, just like Web search. The user is not supported in determining the best items but in identifying those that are *good enough*. (That there are advantages to the second approach are clear, as well as to the first; I do not intend to endorse either in this essay, but merely to demonstrate the extent of their differences.)

Although the adoption of features enabled via new technology has been extensively discussed as a means for genre adaptation — the classic historical studies of Bazerman and Yates being two examples — this process is more typically described as the manipulation of new technologies by writers to further their rhetorical goals, as when Spinuzzi's SEO analysts use templating features of Microsoft Word to facilitate the customization of report elements. In this case, however, the technological features that produce ultimate genre change, in terms of perceived use function of metadata systems, are deployed independently from the "writers" and the generation of primary content. Moreover, the initial effects of a keyword search mechanism are compounded through aggregation, in a manner similar to that described for the subject-heading case study. The difference in scale between 364 results for a catalog-only author search and 632 results for a catalog-only keyword search is significant, but not tremendous. However, the difference between 364 results and the 65,439 results of keyword searching for the integrated discovery system is transformative. This leap in collection scale not only solidifies the keyword search as primary access mode, it makes the previous catalog function, of producing logically structured, individually interpretable results sets, effectively invisible. Just as with the subject-heading case, I suggest that "what users can accomplish by using a digital library" has changed, and, accordingly, the genre of digital library (or catalog) is likewise changed as well.

3.5. Conclusion

In this essay, I have shown how genres can change even within the regulated environments of information systems. In contrast to classic studies of genre development, these changes do not result from the purposeful actions of individual writers, as they respond to changing conditions or take advantage of new technologies. These genre changes result from the aggregation of many apparently inconsequential actions as the regulatory environment remains apparently stable. Sometimes, the shifts in document function that emerge from these aggregated decisions result from changes in document content, as with the case study of *Uncle Tom's Cabin*. Sometimes, however, these functional shifts result from changes in access mechanisms or collection composition that are not conventionally identified as content-related, as the discussion of keyword search and integrated discovery tools demonstrates.

The notion of "writerless" genre change as exemplified through the situation of information systems is, I submit, useful in providing another case where activities previously kept separate — such as the creation of textual content and the design of its material carrier — are shown to be much more inextricably entangled in the digital environment. Johanna Drucker, a scholar of both textual studies and visual book arts, notes how odd it would seem for a traditional literary critic to consider the typeface of a particular edition when interpreting it. As Drucker describes the prevailing approach to literary interpretation, graphic elements may be present in every text, but they are seen as mere distractions to the "real" content: "Material presentation is a necessary, maybe even interesting element of a work, but once we get serious we just 'read'" (Drucker, 2006, p. 267). In the digital environment, scholars like Drucker contend, the meaningful relationship between content and presentation becomes increasingly apparent. In addition to content and presentation, this essay suggests that, especially in the context of unpredictable aggregation, the relationship between "automated" or "regulated" content and the means of its generation (whether via standards or algorithmic production) must also be more carefully considered.

References

Andersen, J. (2006). The public sphere and discursive activities: Information literacy as sociopolitical skills. *Journal of Documentation*, *62*(2), 213–228.

Andersen, J. (2008). The concept of genre in information studies. *Annual Review of Information Science and Technology*, *42*, 339–366.

Bawarshi, A. (2003). *Genre and the invention of the writer*. Logan, UT: Utah State University Press.

Bazerman, C. (1988). *Shaping written knowledge: The genre and activity of the experimental article in science*. Madison, WI: University of Wisconsin Press.

Bhatia, V. (2004). *Worlds of written discourse: A genre-based view*. London: Continuum.

Buckland, M. (2012). Obsolescence in subject description. *Journal of Documentation*, *68*(2), 154–161.

Coyle, C., & Hillmann, D. (2007). Resource Description and Access (RDA): Cataloging rules for the 20th century. *D-Lib*, *13*(1–2). Retrieved from http://www.dlib.org/dlib/january07/coyle/01coyle.html

Drucker, J. (2006). Graphical readings and the visual aesthetics of textuality. *Text*, 16, 267–276.

Gilliland, A. (2008). Setting the stage. In M. Baca (Ed.), *Introduction to metadata (Version 3)*. Los Angeles: Getty Research Institute. Retrieved from http://getty.edu/research/publications/electronic_publications/intrometadata/

Hoffman, G. (2009). Meeting users' needs in cataloging: What is the right thing to do? *Cataloging and Classification Quarterly*, *47*(7), 631–641.

Miller, C. (1984). Genre as social action. *Quarterly Journal of Speech*, *70*, 151–167.

Rose, M. (2012). The ship has sailed and we aren't on it: How catalogers could support user tasks and why we won't. *Journal of Library Metadata*, *12*(2–3), 127–139.

Segal, D. (2011). The dirty little secrets of search. *The New York Times*, February 12, 2011.

Spinuzzi, C. (2004). Four ways to investigate assemblages of texts: Genre sets, systems, repertoires, and ecologies. *Proceedings of SIGDOC*, pp. 110–116.

Spinuzzi, C. (2010). Secret sauce and snake oil: Writing monthly reports in a highly contingent environment. *Written Communication*, *27*(4), 363–409.

Spinuzzi, C. (2011). Genres and generic labor. In C. Bazerman, C. Dean, J. Early, K. Lunsford, S. Null, P. Rogers, & A. Stansell, (Eds.), *International advances in writing research: Cultures, places, measures* (pp. 487–505). Fort Collins, CO: The WAC Clearinghouse and Parlor Press. Retrieved from http://wac.colostate.edu/books/wrab2011/

Tennis, J. (2012). The strange case of eugenics: A subject's ontogeny in a long-lived classification scheme and the question of collocative integrity. *Journal of the American Society for Information Science and Technology*, *63*(7), 1350–1359.

Theimer, S. (2012). A cataloger's resolution to become more creative: How and why. *Cataloging and Classification Quarterly*, *50*(8), 894–902.

Yates, J. (1989). *Control through communication: The rise of system in American management*. Baltimore, MD: Johns Hopkins University Press.