

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Documentary Editing: Journal of the Association
for Documentary Editing (1979-2011)

Documentary Editing, Association for


2011

Rethinking Digital Editing Practices to Better Address Non-Canonical Texts

Amanda A Gailey

University of Nebraska - Lincoln, gailey@unl.edu

Follow this and additional works at: <http://digitalcommons.unl.edu/docedit>

 Part of the [Digital Humanities Commons](#), [Other Arts and Humanities Commons](#), [Reading and Language Commons](#), and the [Technical and Professional Writing Commons](#)

Gailey, Amanda A, "Rethinking Digital Editing Practices to Better Address Non-Canonical Texts" (2011). *Documentary Editing: Journal of the Association for Documentary Editing (1979-2011)*. Paper 12.
<http://digitalcommons.unl.edu/docedit/12>

This Article is brought to you for free and open access by the Documentary Editing, Association for at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in *Documentary Editing: Journal of the Association for Documentary Editing (1979-2011)* by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Rethinking Digital Editing Practices to Better Address Non-Canonical Texts

Amanda Gailey

This article stems from my recent work on *Race and Children's Literature of the Gilded Age* (RCLGA),¹ a digital archive that aims to provide a heavily annotated resource for scholars and students of literature, history, African American studies, visual communication, and education to examine how adults wanted children to think about race during the era of Jim Crow. I edit the archive with Gerald Early, Professor of Modern letters, English, African studies, and African American studies at Washington University in St. Louis, and D. B. Dowd, Professor of Communication Design and American Culture Studies, also at Washington University. When complete, RCLGA will include literature, illustrations, and popular-culture materials featuring characters of different races primarily intended for a juvenile audience between the end of the Civil War and the publication of *The Brownies' Book*, the first American mass-market periodical for minority children, in 1920–1921. In some cases, the authorship of this material is collaborative, corporate, or altogether unknown. What binds the materials together is that they all provide evidence of how popular media marketed to children or families during the period of Jim Crow helped to assert, reinforce, and, occasionally, diminish racial inequity.

The materials seem almost defined by their unsuitability for a scholarly editing project. Virtually all of the materials are uncanonical or decanonized texts; their authorship is frequently slippery or of little interest, for many of the texts are derivative works, sometimes by one or more uncredited authors, and their afterlives in unauthorized or appropriated forms are often more significant to our study than their pristine origins. The conventional ways of conceptualizing a scholarly edition or digital archive, along with the methods and technologies developed around conventional editions, while eminently reasonable for certain types of materials, have proven unwieldy and inappropriate for ours.

¹ This is a working title and will likely change before we make the archive public.

In this article I would like to discuss how we should treat literature that falls into the chasm between scholar-led digital editing, usually organized around one canonical author or text, and mass digitization projects, which lightly treat large numbers of texts with little guidance or claims about their literary or historical value. Neither model offers much support for inquiries about, say, how one text influenced another or how generic or thematic similarities stretch across works by different authors in even the same time period. I will suggest that emerging semantic Web technologies, combined with existing digital markup practices, may be the way to accommodate a wider variety of interest in many literary texts.

Author- and single-text-centered editions have long been important tools for literary scholarship, and it would be surprising if they lacked digital equivalents. However, the study of some kinds of literature is not best accomplished through single-author or single-work editions, yet still benefits from the structure and editorial attention of a rigorously edited and deeply marked-up project. The works of Joel Chandler Harris, which I have been editing for RCLGA, serves as a case in point.

If editorial work moved faster than glaciers, there would likely be a sprawling, multi-volume print edition of Joel Chandler Harris's works in every American research library. In the mid-1950s, when editors schooled in the latest techniques of professional editing undertook the preparation of modern editions of so many American authors, Harris probably seemed like a prime candidate for such work. At the time of his death in 1908 he was second in popularity only to his admirer Mark Twain, and Theodore Roosevelt published a letter mourning the loss of a national treasure, declaring Harris's fiction the most likely of American works to endure.² In the 1920s, over a decade after Harris's death, a survey of U.S. high school and college teachers showed that Harris was considered one of the five most important authors in the United States.

Harris published dozens of novels and collections of short stories over his literary career, which stretched from 1881 to 1908, but his most popular works were his Uncle Remus books, in which a loyal former slave tells folk stories to an unnamed white child. Today, Harris's phonetic spellings of nineteenth-century middle-Georgian African-American dialect, which strike many readers as difficult or offensive, and his paternalistic approach to Remus and occasional

² For information on Harris's popularity, see Walter Brasch, *Br'er Rabbit, Uncle Remus, and the 'Cornyfield Journalist': The Tale of Joel Chandler Harris* (Macon: Mercer University Press, 2000). In a letter published in *Uncle Remus's Magazine* (September 1908, p. 5) Roosevelt opined, "I very firmly believe that his writings will last; that they will be read as long as anything written in our language during his time is read."

implicit nostalgia for the antebellum days, have helped boot him from the canon. Perhaps most damaging to Harris's reputation, however, was the appearance of *Song of the South*. Produced by Walt Disney and released by RKO Radio Pictures in 1947, the film trades on the worst aspects of the Harris tales: the more complex Remus character of Harris's books is caricatured into a hyperbolically loyal and happy servant to white children, and today, over sixty years after its premiere, Disney views the film as an embarrassment and refuses to re-release it.

Though Harris himself has fallen into obscurity, and though we no longer view his works as unproblematically good, his texts and their reception suggest interesting and instructive patterns about late nineteenth-century American attitudes toward race and culture. To study these patterns, though, requires a perspective very different from the view afforded by an author- or work-centered edition. I started working with Harris's texts almost three years ago and began by scanning, transcribing, proofing, and encoding the first editions of his Uncle Remus books. Two years ago, when Emory University in Atlanta agreed to let RCLGA use their Harris holdings, I approached their vast collection from the perspective of an author-centered archive. As I planned how to use limited time to go through thousands of special-collections items, the best approach seemed to be to concentrate on the early drafts of his work. Indeed, the materials for an author-centered digital edition of Harris's work are ripe for the picking. Given enough time in Emory's special collections, we could trace many of his tales from their first drafts through first or final publication. But such an approach would have a hard time answering the question, "Who cares?" Harris's texts are entertaining reads and fascinating glimpses into U.S. racial history, but certainly the best treatment of them is not presenting Harris as he may have once been viewed: a highly canonical genius whose compositional process elicits scholarly curiosity or admiration. It is the reception of Harris's works, how they were pirated, appropriated into popular culture, and generally disseminated into American racial consciousness that is of interest, and studying these is not in the least enabled by an author-centered edition.

Our critical interest in Harris's Uncle Remus tales is in many ways similar to ongoing scholarship on Harriet Beecher Stowe's *Uncle Tom's Cabin*, which in recent years has enjoyed attention from scholars who examine its cultural significance through the many sympathetic and hostile appropriations it spurred: parodies, homages, minstrel shows, and so on. Stowe's and Harris's central characters followed a very similar path through American and international culture. Both Uncle Tom and Uncle Remus were born in the pages of American periodicals—Stowe's in the *National Era* and Harris's in the *Atlanta Journal Constitution*. Both Harris and Stowe wrote well-intended but sentimental and paternalistic depictions of black characters in an effort to effect social change.

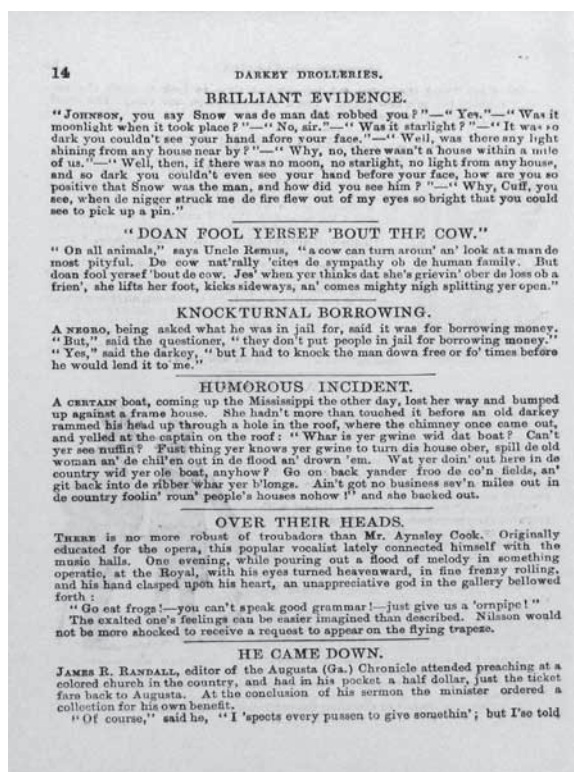


Figure 1: Excerpt from *Darkey Drolleries* by "Uncle Remus." London: John and Robert Maxwell, 1883. Courtesy of Manuscript, Archives, and Rare Book Library, Emory University.

Stowe, of course, sought the end of slavery, and Harris, writing at the height of Jim Crow in the American South, hoped to humanize African Americans to his white readers in an effort to end the epidemic of lynchings. Stowe's and Harris's political goals, however flawed they now may seem in execution, were progressive in their times and elicited both admiration and hostility.

Stowe's and Harris's moral earnestness made them prime victims of parodists and hacks. Just as Stowe's characters were quickly subsumed into consumer culture and the minstrel stage, Uncle Remus was featured in pirated publications, abridgments, household decorations, advertisements, and corporate logos.

A closer look at a derivative British publication, *Darkey Drolleries*, will help demonstrate a problem with digital editing (see Figure 1). This booklet, published in London in 1883 by John and Robert Maxwell, purports to be by Uncle Remus, but is in effect a printed minstrel show much like the stage adaptations of *Uncle Tom's Cabin*, in which

Stowe's originally progressive characters were conscripted to racist aims. Amid passages actually written by Harris, the publishers inserted materials they either wrote or recycled from other publications, much of which is far more vehemently racist than anything Harris ever wrote. So although the entire publication is ascribed to Uncle Remus, this page consists of a clip from a British compilation of American humor, a reflection on cows attributed to Uncle Remus, an excerpt from a book of humor published in Pittsburgh, a racist anecdote circulated on American postcards, an anecdote about a British opera singer, and a story reprinted from a Georgia newspaper. Of the six pieces on this page, the only one for which we cannot locate a source is the one spoken here by Uncle Remus—it does not appear to have come from anything Harris wrote. The whole booklet, predictably, is illustrated with racist caricatures that could be stock drawings or may have been taken from other publications.

Darkey Drolleries is evidence of how literary piracy influenced the proliferation of racist imagery. Considered alongside the postcards, menus, coloring books, and toy sets that all bore Harris's characters, it seems that

the cultural reach of Harris's characters far exceeded his grasp. It is this reach that is of most interest about Harris, but it is the author's grasp that defines most editorial undertakings. Unfortunately, some of the most provocative and interesting materials relating to Harris fall into a no-man's land that seems beyond the scope of a single-author edition but would not be adequately noted or otherwise made available in mass digitization efforts.

One of the few digital projects to trace the cultural reconfigurations of an American literary text is Stephen Railton's *Uncle Tom's Cabin in American Culture*,³ which is full of fascinating material and editorial insights but beset by technological and navigational problems. Railton's site includes images of several editions of the book; Stowe's own *Key to Uncle Tom's Cabin* and adaptation of the novel for the stage; "pretexts" that illustrate the culture into which Stowe was publishing her work; numerous reviews of the book; adaptations of the book for children; 3D manipulable images of memorabilia based on *Uncle Tom's Cabin*; and images of the book's eventual transformation for stage and screen. All of this traces the text's trajectory from its roots in mid-nineteenth-century abolitionism to its transformation into a twentieth-century industry of racial degradation and caricature.

The problem with the site, which is not unrelated to its inconsistent interface and difficult navigation, is that few developed methods available to digital literary scholarship support this kind of approach to texts. For example, Railton is rightfully interested in the covers of early editions of *Uncle Tom's Cabin*. From the golden age of conspicuous literary consumption, these covers speak to what publishers and purchasers found important and beautiful about the text. They worked as the marketable face of the book, and if the ubiquitous still-uncut pages of nineteenth-century gift editions tell us anything, it is that the face of the book most frequently held the owner's interest. Yet TEI (Text Encoding Initiative), the *de facto* encoding standard for digital editing projects, does not even have a single tag, much less a developed module, for describing the outside of a book—something that is relevant to many literary projects.⁴ The *Walt*

³ See <http://utc.iath.virginia.edu>.

⁴ Both scholarly digital editions and mass digitization projects are built using the guidelines published by the Text Encoding Initiative (TEI), the *de facto* standard for digital editing in the humanities. TEI provides a vocabulary of several hundred terms that editors may use to label structural features and significant content in the transcription of a text within an XML (Extensible Markup Language) file. TEI provides a common language, allowing editors from different projects to communicate about their materials and sometimes even aggregate them. It has proven invaluable to the development of digital editing, and the fact that editors can customize it—selecting and tailoring the TEI terms or "tags" that best work for their projects—makes it a tool that provides not only a common language for the digital editing community, but one that is pliant enough to serve an array of editorial interests.

Whitman Archive,⁵ for example, understandably omits cover information from its encoding, since it is unsupported by TEI. However, the 1856 edition of *Leaves of Grass* was widely known at the time for the adulatory Emerson quote—“I greet you at the beginning of a great career”—that Whitman brazenly reprinted on the spine. If the *Archive* wanted to include this they would have to develop an ad-hoc TEI extension. One of RCLGA’s customizations to TEI was to create a simple tag for the illustrator of a book, a person as important to children’s literature as the author. The orthodox TEI approach to treating illustrators is, bizarrely, as a specialized kind of editing—that is, TEI recommends encoding the illustrator this way

<editor role=“illustrator”>A. B. Frost</editor>

as though the illustrator were merely another corrupting or altering influence on a pristine text.

The lack of an <illustrator> tag in TEI reflects a bias in the way the vocabulary allows projects to describe texts. TEI works best for digital editions that view a single and singly intended text as the fundamental unit of the archive. The literary structures of that text, but not its bibliographic structures, its relationship to other texts, nor its collaborative aspects, are robustly supported by the TEI tagset. A project designed around a core of stand-alone texts written by one author is much more suited to TEI than a thematically oriented project or one that examines textual transmission and appropriation. In the case of the author-centered archive, the design of the project matches up with the nesting structures of TEI: the identity of the author contains individual texts which are comprised of chapters, which hold paragraphs, and so on. But if a project hopes to examine the kinds of connections and cultural dispersions that Railton’s work addresses, for example, TEI may begin to seem like a hindrance: a significant investment of time and labor into tagging that supports little of the intellectual interest of the project. Given that the major U.S. funding agencies all but state a requirement of TEI compliance for digital editorial projects,⁶ the author-centered model is not only implicitly encouraged by the current granting system but is also clearly the path of least resistance for anyone with an interest in digitally editing American literature.⁷ For many projects, though, it seems that we lack good

⁵ See <http://www.whitmanarchive.org>.

⁶ See, for example, the guidelines for Scholarly Editions grants from the National Endowment for the Humanities: <http://www.neh.gov/grants/guidelines/editions.html>.

⁷ Ann Gordon discusses how grants for editing historical papers have skewed the selection of subjects in “Experiencing Women’s History as a Documentary Editor,” *Documentary Editing*, 31 (2010): 1–9.

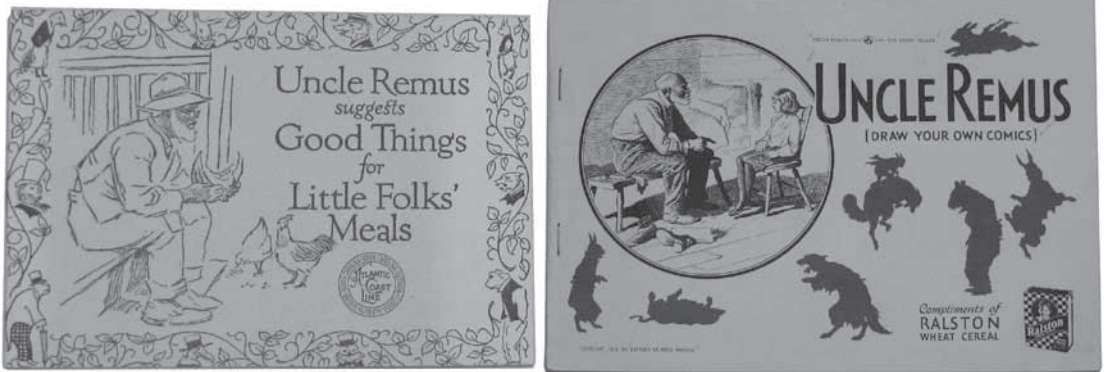


Figure 2: Uncle Remus ephemera from Atlantic Coast Line and Ralston Wheat Cereal. Courtesy of Manuscript, Archives, and Rare Book Library, Emory University.

editorial standards for describing what is most of value: how, in Joseph Grigely's words, those "post-textual reconfigurations of a work tell us something about the personality of a culture."⁸

TEI is indispensable for many digital editing tasks, but was simply not intended to note the relationships among ephemera such as a children's menu, a do-it-yourself comic, or the other many merchandise tie-ins and other cultural goods generated by the publication of the Uncle Remus tales (see Figure 2). Similarly, it is not suited to capture what is of interest in the array of materials pertaining to *Uncle Tom's Cabin* that Railton has collected, or the larger patterns of co-opting Uncle Tom and Uncle Remus, which ranged from friendly retellings to degrading parodies. We have a markup vocabulary for noting intricate structures within a single text, but lack a graceful way of noting patterns or relationships to which a text belongs.

As a case in point, I would like to look at Harris's first story collection, *Uncle Remus: His Songs and Sayings*, from 1880. This collection was arguably his most influential, and its story about Brer Rabbit's encounter with a tar baby would be Harris's most widely recognizable tale. RCLGA has encountered a variety of materials, both texts and nontextual objects, that were directly or indirectly based on Harris's Uncle Remus tales and that demonstrate the tales' cultural influence. In some cases an item was clearly influenced by a particular book or story, as with "Tar-Baby Nails." Other items make use of characters that recur in several of his collections.

⁸ Joseph Grigely, *Textuality: Art, Theory, and Textual Criticism* (Ann Arbor: University of Michigan Press, 1996): 46.

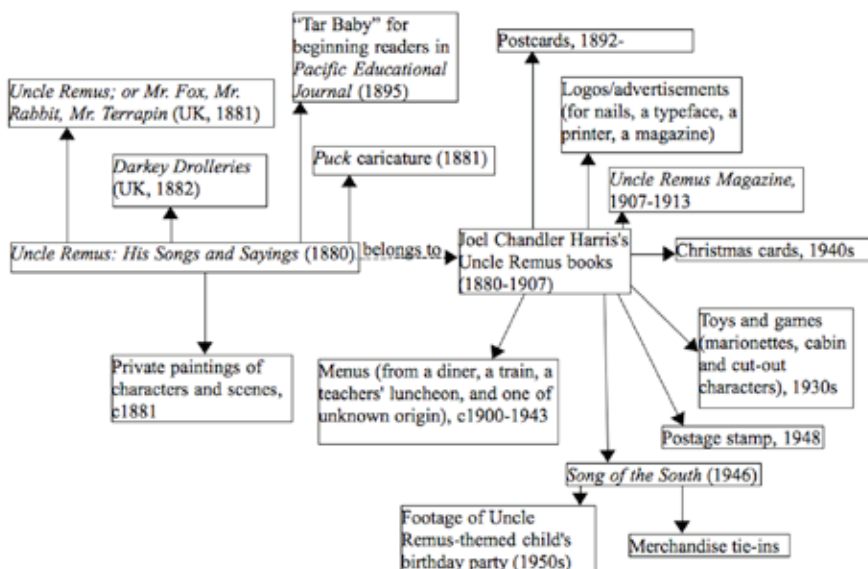


Figure 3: A schematic illustrating how Harris’s Uncle Remus books inspired derivative cultural works.

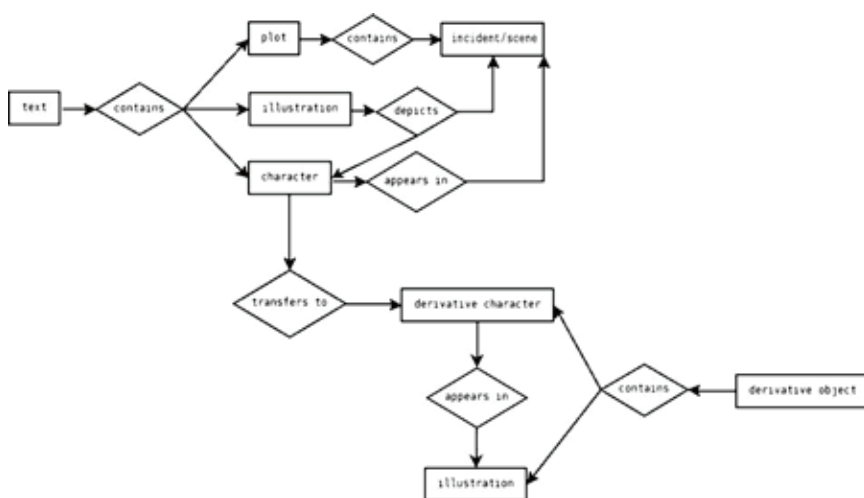


Figure 4: A model of how a textual component is appropriated by another work.

Figure 3 illustrates the relationships among many materials we have encountered while working on Harris's texts. For the sake of simplicity the diagram omits materials commonly encountered in author-centered archives, such as manuscript drafts. I have separated *Uncle Remus: His Songs and Sayings* from the rest of the Remus texts in order to illustrate that many derivatives directly relate to one text, while others, produced after many of the books were published and making use of recurring characters, cannot be traceable to a particular text. Except for the dotted arrow noting the special relationship between the 1880 text and its parent category, the arrows in the diagram indicate the direction of demonstrable influence.

Viewed this way, it is easy to see how this now fairly obscure text functioned as a cultural vector. However, this diagram does not capture many details of the relationships that are of interest to literary scholarship. A few details that would likely be of wide interest include:

1. What kind of thing is the appropriation?
2. Was the appropriation authorized by Harris?
3. Is the appropriation sympathetic to or critical of the aims of the source text?
4. What aspects of the source text were appropriated: characters, illustrations, plot, direct language?

There are more questions we would want to ask, of course, but these few may serve as examples. A model that could support these kinds of queries would need to be ontologically nuanced; that is, it would need to be a carefully constructed formal system that describes the entities, their properties, and their relationships to each other. Figure 4 attempts to represent how a character in one object (a book) is appropriated into an illustration in another object (a menu). Because space is an issue I have kept the figure simple and only included here some representative entities and attributes.

A few technologies could allow us to implement this model. Relational databases appeal to some literary archives that attempt to foreground relational structures.⁹ Recently, the viability of relational databases as a tool for modeling

⁹ See, for example, Ed Whitley's *Vault at Pfaff's* (<http://digital.lib.lehigh.edu/pfaffs/>), a study of the bohemian community in New York that served as a nursery for several important mid-nineteenth century American writers. Whitley realized early in his work that the relationships among these authors and the many texts they published in a particular newspaper were of much more scholarly interest than the finely tuned editing of any particular text. He and a colleague at Lehigh University's library created a database that would allow them to express these relationships and store an impressive quantity of annotations on the writers (Whitley, 5–6). Railton's site on *Uncle Tom's Cabin* would benefit from a relational database, but when the site was updated to conform to technological standards, the focus was on migrating the data into TEI-compliant XML.

texts has been called into question, most notably in a lively *PMLA* exchange between Ed Folsom, co-editor of the Walt Whitman Archive, and several respondents, including Jerome McGann. McGann makes a compelling point about the strength of markup and limitations of databases in literary editing:

For scholars interested in migrating our cultural inheritance to digital environments, databases are by no means the most useful tools for the task.

. . . The inline markup approach of the Text Encoding Initiative . . . became a standard for digitizing literary works for a reason. . . . Let's be clear. The TEI and XML do not adequately address the problem of knowledge representation that is the core issue here—that is, how do we design and build digital simulations that meet our needs for studying works like Whitman's—but they get a lot further along with that task than do database models. They are better because they model some of the key forms of order that are already embedded in textual works. . . . They are better because they understand that works like poems and novels are already marked data.¹⁰

I believe the conflict between inline markup and relational models that McGann addresses here is a false one. McGann is correct about the suitability of markup to the digitization of individual texts, for which databases would be an ill-fitted tool. But here he seems to conflate individual texts with “our cultural inheritance,” which really is (as he later acknowledges) more than the sum of its parts. Markup is the best tool for those parts, but for representing the complex relationships among texts we need a technology that is suited to describing relationships. McGann and others in the *PMLA* exchange seem to use “database” to mean a non-narrative representation of ontologically discrete objects and their attributes and relationships to one another. However, McGann's criticism of database as a form of knowledge representation and accommodation seems to arise from occasionally conflating this abstract definition with the particular technologies used to build relational databases, which are much more rigid and limiting than database in the abstract. He explains that “databases and all digital instruments require the most severe kinds of categorical forms. The power of database—of digital instruments in general—rests in its ability to draw sharp, disambiguated distinctions.”¹¹ He argues that card catalogs are more amenable

¹⁰ Jerome McGann, “Database, Interface, and Archival Fever,” *PMLA*, 122 (2007): 1589.

¹¹ *Ibid.*, 1590.

to literary research than digital databases because their physicality allows people to “intervene” when strict categorizations are not helpful—by jotting notes, cross-references, and other exceptional information on the cards. He concludes that databases ultimately fail to accommodate and contain our knowledge because scholars have multiple and ever-changing interests in texts and their production and reception histories. As he puts it, “Scholars do not edit or study self-identical texts. They reconstruct a complex documentary record of textual makings and remakings, in which their own scholarly investments directly participate.”¹² Crucially, though, this rigidity may be inherent in SQL (Structured Query Language) based databases, but it is not inherent in other ontological technologies that identify objects and declare relationships, such as new semantic web technologies.

McGann’s critique of database thus seems at least twofold: first, inline, marked data better captures textual structure and content than database, and second, the strict categorization required by databases defies the diverse, amorphous, and ongoing record of our engagement with texts. The first complaint seems accurate but ultimately irrelevant, since treating an individual text with inline markup is not incompatible with treating the text as an object in a database that records inter- or extra-textual information. The second complaint seems targeted at a specific relational database technology, and not at the more abstract idea of a model that records information about texts, their attributes, and their relationships.

Semantic web technologies such as RDF (Resource Description Framework) and OWL (Web Ontology Language) are proving to be viable, flexible alternatives to relational databases, and can accommodate an expanding and diverse set of claims about entities, their properties and relationships. While admittedly difficult to learn and technically implement in comparison to inline markup or the relational database technologies that McGann finds too limited, the formal specifications of the semantic web, expressible in XML, allow projects to specify particular entities and relationships among them within sophisticated ontologies. Semantic web technologies provide ontological rules and a syntax for expressing them. For example, if we were to describe a set of relationships involved in textual transference in OWL, we could claim the following:

- 1) Uncle Remus is a character in “The Wonderful Tar Baby Story”
- 2) Joel Chandler Harris wrote “The Wonderful Tar Baby Story”
- 3) A toy advertisement features Uncle Remus.

¹² *Ibid.*, 1592.

The inferential rules we can describe through OWL would later allow us to derive from this information that this toy advertisement features a character that Joel Chandler Harris created, even though we never directly stated that. This is a simple example, but because the technologies allow us to provide the ontology with new entities continually, it can grow complex very quickly and ultimately allow us to derive sophisticated information about two entities inferred from a multitude of single, separately entered statements about them. Essentially, the data-interchange standards of the semantic web are expansible in ways that can satisfy McGann's and others' understandable discomfort about the brittleness of relational databases.

Already some digital humanists are exploring how semantic web technologies such as RDF can complement inline markup. For example, recently in *Literary and Linguistic Computing* Ariana Ciula and Tamara Lopez explained how the Henry III Fine Rolls Project uses RDF and OWL to help express relationships among historical figures.¹³ Other projects, such as NINES (Networked Interface for Nineteenth-Century Electronic Scholarship), use RDF to express orthodox metadata about individual texts.

Semantic web technologies, combined with a constrained vocabulary tailored to the purpose, could help literary scholars and cultural historians track the dissemination of characters and images through texts and other artifacts. Further, the flexibility and expansibility of semantic web technologies such as RDF and OWL could allow for aggregation: individual projects describing historically or thematically similar materials could combine records and allow for the mapping or graphing of relationships among materials across projects. But currently, the lack of a developed method for tracking cultural transmission is a significant lacuna in digital literary scholarship. As we seek to build upon the rich tradition of the collected edition, an ontological framework for describing intertextual relationships could prove fruitful.

¹³ "Reflecting on a Dual Publication: Henry III Fine Rolls Print and Web," *Literary and Linguistic Computing*, 24 (2009): 129–41.