Towards a Rationale of Audio-Text

Tanya E. Clement <tclement_at_ischool_dot_utexas_dot_edu>, University of Texas at Austin

Abstract

Digital humanities scholars have made a tradition of problematizing our understanding of textuality through discussions concerning the design of information systems for texts that, in many cases, still look like books. This discussion is concerned with how creating opportunities for studying audio texts further complicates our understanding of "the rationale of a textualized document," defined by Jerome McGann as "the dynamic structure of a document as it is realized in determinate (artisanal) and determinable (reflective) ways". This discussion frames a rationale of audio text within the context of developing information infrastructures for accessing audio texts. I introduce a tool called ARLO that we have been developing in the High Performance Sound Technologies for Access and Scholarship (HiPSTAS) project (http://www.hipstas.org) for accessing and analyzing sound collections alongside new standards being proposed for the development of audio visual (AV) metadata and content models. The discussion concludes by considering what these interventions tell us about how a rationale of audio textuality helps us rethink rationales of text in digital environments.

1. Introduction

In the 1980s and 1990s, *A Critique of Modern Textual Criticism*, *The Textual Condition*, and *Bibliography and the Sociology of Texts* laid theoretical groundwork for understanding the social nature of text [McGann 1983] [McGann 1991] [McKenzie 1999]. Namely, McGann and McKenzie argued against the traditional editing practices that included identifying a *definitive text*, a *copy text*, an *ideal text*, an *Ur text*, or a *standard text* for scholarly editions. They reasoned that these practices obfuscated the social histories and contexts that meaning-making with texts must reflect, including how texts are produced, transmitted, received, and consumed. McCann, McKenzie, and others emphasized textual theories that reflected the significance of materiality, versioning, technical transmission, and institutional contexts. McGann's "The Rationale of Hypertext" applies these social text theories in the digital realm [McGann 1995] where an escape from "book-bound" and "fixed point" navigations indicates a new kind of social condition in the digital age of hypertexts. In the fifteen years since, we have seen multiple scholars (such as [Bryant 2002] [Drucker 2002] [Liu 2004] [Kirschenbaum 2008]) take up McGann's charge to use social text theory and the seemingly expanded perspective afforded by the digital environment to rethink what McGann has called "the rationale of a textualized document" [McGann 2001, 137]. Of this tradition, this discussion is concerned with how digital audio texts further shape rationales of texts more generally.[1]

First, after defining some key terms, I frame this problem within conversations about textuality that have arisen around developing information infrastructures for accessing *verbal texts*. What I am calling verbal texts are documents such as books, articles, letters, and audio for which the content has been represented in information systems as linguistic content or words. Second, in order to situate rationales of verbal texts in the context of audio-texts, I describe a specific example – the development of a tool for accessing and analyzing sound collections in the High Performance Sound Technologies for Access and Scholarship (HiPSTAS) project (http://www.hipstas.org). Third, I use this history and this specific example to consider how current content models and metadata standards reflect rationales of audio texts both in digital humanities and beyond. Finally, I conclude by considering what these interventions tell us about how we may rethink a rationale of textuality in digital environments.

2. Rationales of Text

The *rationale* of a text comprises the principles by which we argue that text makes meaning. This discussion is guided by McGann's definition of text's *rationale* as part of a social condition which he contends is made apparent through "the dynamic structure of a document as it is realized in determinate (artisanal) and determinable (reflective) ways" [McGann 2001, 137]. With this definition, McGann is focused primarily on verbal texts, but he is also considering how the physical object of the text is made as artisanal properties. These properties might include how a book is bound or which picture might be on the cover of a particular edition or which font or words were chosen. The text's reflective properties, on the other hand, point to how this physical and conceptual text is understood by a reader who might make meaning about (or reflect on) that book binding, that image, and that font or word choice. These two kinds of information, which McGann calls determinate and determinable, jointly or dynamically comprise the verbal text's meaning in social text theory.

Additionally, by placing special emphasis on how interpretation relies on the realization of a *document,* McGann situates these determinate and determinable properties within the context of processes for textual production, transmission, and reception in social text theory. Unlike the concept of *text*, which can span multiple iterations of a work, *document* is understood as a situated object in time and place: it is "any physical or symbolic sign, preserved or recorded, intended to represent, to reconstruct, or to demonstrate a physical or conceptual phenomenon" (Briet, 1951 quoted in [Buckland 1997]). Much as social text theory is concerned with the publishing and social systems (the scholarly infrastructure) that define textuality, document theory in information science likewise comprises the study of the *document* within particular contexts or systems of organization and distribution called *documentation* including collecting, standardizing, filing, classifying, copying, disseminating, and preserving documents [Buckland 1997, 805]. Historically, these concerns have been shared by many information scholars who consider the *document* in terms of these analog, digital, and decidedly social, processes of documentation (for examples, see [Otlet 1990]; [Frohmann 2004]; [Briet 2006]; [Feinberg 2010]). McGann's rationale of text as "the dynamic structure of a document as it is realized" also refers to these documentation processes (among others) as essential means by which we come to understand textuality or the rationale of text.

Before proceeding, it is important to note that text can be realized in different ways. D.F. McKenzie defines text as "verbal, visual, oral, and numeric data, in the form of maps, prints, and music, of archives of recorded sound, of films, videos, and any computer-stored information, everything in fact from epigraphy to the latest forms of discography" [McKenzie 1999, 13]. Further, in the digital envinronment we can process texts as other texts such as sonifying or rasterizing verbal texts, a consequence of "the digital code [that] renders commensurate texts, images, and sounds" [Enrst 2012, 84]. It is also important to realize that the kinds of text with which McGann and others in digital humanities have been primarily interested are literary, cultural, or aesthetic. As such, these kinds of verbal texts are usually self-reflective or "playfully" aware of their status as realized texts. That is, a literary text such as a poem asks the reader to take into consideration how the poem looks on the page, how the words sound together, how the choice of one word over another or the lack of a word makes meaning. As a result, these kinds of texts have what McGann calls a "generic rationale... to maximize attention to the structure and interplay of the textual orders" – orders such as the linguistic (semantic) and bibliographic (graphical) codes present on a page [McGann 2001, 138]. Literary, cultural, and aesthetic texts often emphasize how blurring the boundaries between these elements [McGann 2001, 137] can help form the text's meaning.

This discussion will address how sound in spoken word audio may be realized in digital information systems. While a wide range of literary, cultural, and aesthetic artifacts may be classified as *text*, the primary focus for definitions (or rationales) of text in digital humanities have focused on the words of verbal texts in exclusion of other meaning making signifiers. Consider McGann's footnote to his definition of *text* in which he admits that phonological and tactile elements (such as how words sound and how a page feels to the touch) are also signifiers but are outside of his "scholarly expertise" [McGann 2001, 254]. Meant to help address this sonic gap in McGann's rationale and in order to reconsider the nature of textual rationales more broadly, the rationale of audio textuality considered here is accomplished within the context of scholars working with literary, aesthetic, and cultural aspects of poetry in recorded performances.

3. Rationales of Digital Text

Many conversations in digital humanities are concerned with how we model literary texts in systems that reproduce, transmit, display, and analyze text. These discussions typically detail the difficulties creating a model of text that seems to both represent the scholar's perceived rationale of that text (as "authentic") with the explicitness and consistency that a computer that can "compute" or process those elements requires [Johansson et al. 1991a].[2] As Willard McCarty reminds us, creating a computationally tractable model of something like "text" is tendentious: it "forc[es] us to confront the radical difference between what we know and what we can specify computationally, leading to the epistemological question *of how we know what we know*" [McCarty 2004]. Floyd and Renear also note that an ontology for modeling textuality through computers requires a balance between recognizable and computable: the model "need not reflect the latest theories of modern physics, but it should nevertheless at least be internally consistent, and as much as possible avoid clashes with commonsense beliefs" [Floyd and Renear 2007].

In order to better explicate these tensions between a tractable text and an authentic representation of text, I outline two models for computing texts that are widely used by the humanities and library communities and have been critiqued in DH for seemingly inauthentic representations of textuality. The first is a content model called Ordered Hierarchy of Content Objects (OHCO), which has a history in typesetting and printing and was the model upon which the TEI (Text Encoding Initiative) Guidelines for Electronic Text Encoding and Interchange [TEI Consortium 2014] was developed. The second example, Functional Requirements for Bibliographic Records (FRBR), is a model meant to facilitate the organization and discovery of cultural artifacts through library catalogs (International Federation of Library Associations).

3.1. Ordered Hierarchy of Content Objects (OHCO)

TEI, one of the first and most widely adopted models of text used in developing information infrastructures in DH was based on the idea of text as an ordered hierarchy of content objects (OHCO) [DeRose et al. 1990]. In the OHCO model, texts are primarily comprised of "linguistic objects" in which form (sections, paragraphs, and quotation blocks) and content (the words that comprise these entities) are separate but related aspects. As a hierarchical model, bigger objects such as chapters or volumes "contain" smaller objects such as paragraphs, which in turn contain other smaller objects such as lists or quotation blocks. This model reflects a concern with procedural practices for verbal texts in typesetting and printing, which entails adding meaningful metadata to a text as meta-instructions for how to format a document [DeRose et al. 1990, 2]. With the OHCO model, computation is facilitated by the structure of these content objects since they can be consistently maintained across different formats or layouts.

Rationales of text provide for an understanding of text by which we can evaluate whether a model such as OHCO adequately represents textuality more generally. Critiques of OHCO are concerned with the many ways in which elements of a dynamic text overlap and are therefore poorly represented by a hierarchical, nested structure. Literary texts such as novels and poems, for example, often comprise multiple, overlapping themes or rhyming schemes that fall across lines or stanzas. Due to these debates about the model's representation of text, the OHCO authors subsequently "retreated from the simple OHCO thesis" and "conclud[ed] that analytical perspectives do seem to exist and do seem to provide fundamental insights into the nature of texts and the methodology of text encoding" [DeRose et al. 1990]. Subsequent refinements to the OHCO model stipulate that while the OHCO model is both practically and empirically valuable to the general community of text consumers, it is perhaps less valuable in the context of literary, cultural, and aesthetic texts where a text can be understood to reflect a rationale of text with a spectrum of determined and determinable dynamics.

3.2 Functional Requirements for Bibliographic Records (FRBR)

Meant to serve as a model for standards, FRBR was developed primarily as a model or vocabulary for library catalogers and system designers by the International Federation of Library Associations and Institutions (IFLA) in order to standardize practices for finding, identifying, selecting, and obtaining cultural artifacts from libraries [Tillet 2004, 2]. In an attempt to standardize terminologies for abstract concepts, FRBR includes rules for the specific use of general terms including *work*, *expression*, *manifestation*, and *item*. In FRBR, *work* is described as "the story being told in the book, the ideas in a person's head for a book;" *expression* is described as a particular translation or edition (scholarly or popular for example) of the work; *manifestation* is described as a particular publication of the work; and *item* is described as the

"physical object that has paper pages and a binding and can sometimes be used to prop open a door or hold up a table leg" [Tillet 2004, 2]. In its focus on conceptual entity-relationships, FRBR's entities reflect an attempt to disambiguate terms by grouping entities by their characteristics as well as by grouping them according to the relationships that tie them together. Thus, multiple editions or *expressions* of Hamlet can be associated with the same *work*, making all of the editions findable, identifiable, and obtainable according to that relationship.

Though FRBR entities are imagined as "universals", Renear and Dubin show how rationales of text that include textual dynamics at play provide an opportunity for problematizing how the FRBR entities represent texts [Renear and Durbin 2007]. In particular, the authors use the example of the author of a TEI XML (eXtensible Markup Language) document. They discuss a TEI XML document as a complex object that can be simultaneously considered a *manifestation* (or a new publication of the same content) and an *expression* (in the event of a scholarly edition that expresses a version of the text that has not been articulated in this way previously). Due to the kinds of ambiguities these examples represent, Renear and Durbin argue that three of the four FRBR entities in Group #1 are problematic since they are not self identical or unique. They argue that the entities are, in fact, *roles* rather than *types*. That is, while a *type* remains a "rigid" property regardless of the context, a *role* "is brought about by contingent social circumstances" and seems to represent more readily the behaviors that Renear and Durbin observe.

Consequent to making a distinction between *role* and *type* for the FRBR entities, Renear and Durbin refactor the FRBR types in order to suggest new, more "rigid" or distinct types. Adopting John Searle's distinction between natural objects, which are presumed independent of context, and social objects, which are contingent on the social dimension of knowing [Searle 1995], Renear and Durbin call these more distinct types *symbol sequences* (which play the role of realizing a work), *physical kinds* (which play the role of embodying expressions), and *physical objects* (which play the role of exemplifying manifestations) [Renear and Durbin 2007]. While Renear and Durbin admit to their suggestion's inconsistencies, the important point here is the extent to which rationales of text are being used both to problematize as well as reimagine how we can develop rigorous text processing systems for the reproduction, transmission, dissemination, and analysis of texts that are both tractable by computers and authentic to humans.

4. Rationales of audio text

A consideration for rationales of *audio* text as they are realized through systems for processing audio is also productive, especially since, in the humanities, such systems are often developed for literary, cultural, and aesthetic texts such as oral histories, poetry performances, speeches, and storytelling. In order to better understand what is at stake in terms of rationales for audio text, the following describes the rationale of audio text as it is currently being realized in information systems from three perspectives: (1) through the development of a prototype for sound analysis software in the High Performance Sound Technologies for Access and Scholarship (HiPSTAS) project; (2) as a content model represented by the TEI; and (3) in terms of a library cataloging model for describing audiovisual resources through the BIBFRAME Initiative.

4.1 Audio textuality in HiPSTAS

A joint project of the School of Information (iSchool) at the University of Texas at Austin (UT) and the Illinois Informatics Institute (I3) at the University of Illinois at Urbana-Champaign (UIUC), the High Performance Sound Technologies for Access and Scholarship (HiPSTAS) project is developing the ARLO (Adaptive Recognition with Layered Optimization) software. ARLO is a web-based, machine-learning application originally developed for acoustic studies in animal behavior and ecology [Enstrom 1993]. In ARLO, the audio text is represented as a spectrogram. Designed to model a bank of hairs in the inner ear and to vibrate at different audio frequencies in response to sound waves, ARLO creates spectrograms by monitoring and then sampling the instantaneous energy of these "hairs" or tuning forks[3] and then using this data to create a 2D matrix of values (frequency vs. time). These matrices or spectrograms (see example in Figure 1) show a map of sonic energy across time with each row of pixels representing a frequency band and the color of each pixel representing the numerical value of total energy of that particular frequency (or how much the tuning fork trembles) for that point in time. ARLO uses these spectrograms to model sonic features for machine learning processes including clustering (unsupervised learning) as well as classification (prediction or supervised learning).

HiPSTAS was originally funded by the National Endowment for the Humanities as an Institute for Advanced Topics in the Digital Humanities between May 2013 and May 2014 for scholars who are interested in using machine-learning software to analyze audio texts of primarily verbal events such as poetry performances, speeches, and storytelling activities. Many of the HiPSTAS participants used ARLO to analyze poetry performances in PennSound, a web-based archive launched by Charles Bernstein and Al Filreis in 2005 as a noncommercial offering of approximately 30,000 downloadable MP3s — mostly as song-length singles. The PennSound recordings are already retrievable both from a library catalog by authors' names and via Web search engines, but the HiPSTAS participants were primarily interested in analyzing "vocal gestures" in the performances. These gestures such as "the cluster of rhythm and tempo (including word duration), the cluster of pitch and intonation (including amplitude), timbre, and accent" [Bernstein 2011, 126], are always dynamically at play and are especially significant for interpreting poetry performances.

Poets and scholars describe on online HiPSTAS project pages how searching PennSound "sonically" with ARLO allow them to consider new research questions about the performances including a focus on sonic "para-content" such as pitch and laughter. One participant writes, for example, "I have observed from small-scale analysis that timbre and sound duration are indicators of a stressed syllable. Pitch intensity also correlates with stressed syllables, although not in every instance — I would like to investigate further how much pitch intensity correlates with metrical stress" [Boruszak 2014]. Another participant wants to focus on the material artifacts of the recording process, what he calls "para-content audio data" for certain audio texts in order to locate other recordings in PennSound that were recorded as part of a series; he wants "to confirm the suspected provenance of some recordings and to start to bring the recordings together for the first time, perhaps, ever" [Mustazza 2014]. A third participant is interested in moments of laughter that "reveal the presence of the audience, emphasize the construction of sounded poems as a the [sic] product of a dialogue between audience and poet, and change significantly the nature of the poem in question"; focused on "close listenings" of four versions of William Carlos Williams's "This is Just to Say," this participant argues "that a version in which Williams seemed to try to get the audience to laugh and failed showcases the power of laughter at poetry readings, transforming the poem from a whimsical delight to a highly serious, academic reflection on the nature of art and poetry" [Rettberg 2014]. Using ARLO's spectrograms and machine learning processes — which seemed to offer access to these dynamic aspects of audio textuality — the participants began to articulate how these rationales for audio texts might be realized in a system like ARLO.

While provocative, systems like ARLO also provide an opportunity for interrogating whether or not they adequately represent both authentic and tractable texts. For instance, identifying sonic patterns in a system such as ARLO is a time-consuming task of finding and labeling training data that adhere to the listener's understanding of the audio-text's rationale or dynamic structures. Using these human-generated training examples, ARLO creates a model for labeling new, "unseen" audio examples. In the case of a scholar working with recordings, this could mean marking up a sound recording according to some of the sonic aspects the participants identified above such as variances in pitch, recording noises, or the presence of laughter or applause. For example, Figure 1 shows how one could mark up different voices on spectrogram such as "William Owens" (at point "A") and "Robert Frost" (at point "C") as well as silences (at point "B") in order to teach the machine to model these patterns and to identify or classify more of these sonic patterns of interest in a collection.

Labeling examples for machine-learning is ultimately an ontological exercise that requires some standardization. Standardization allows for consistency across machine learning examples as well as results that can be more easily interpreted in terms of a particular discourse community. For instance, instead of compiling their own list of labels, the PennSound scholars chose to use terms provided in the "Transcriptions of Speech" section of the TEI P5 Guidelines because they wanted a set of terms that had been standardized by a community of scholars for whom "mark up" guidelines were the result of sustained study and a concern with representing dynamic rationales of text. Indeed, the elements or terms in the "Transcriptions of Speech" guidelines, which had been vetted by peers who were also versed in both verbal transcription and textual theory seemed to adequately describe the "vocal gestures" also described by Bernstein including *Tempo*, *Rhythm*, *Loudness*, *Pitch*, and *Tension* as well as vocal quality attributes such as *Whisper*, *Breathy*, *Husky*, *Creaky*, *Falsetto*, *Resonant*, *Unvoiced Laugh or Giggle*, *Voiced Laugh*, *Tremulous*, *Sobbing*, *Yawning*, and *Sighing* [TEI Consortium 2014].

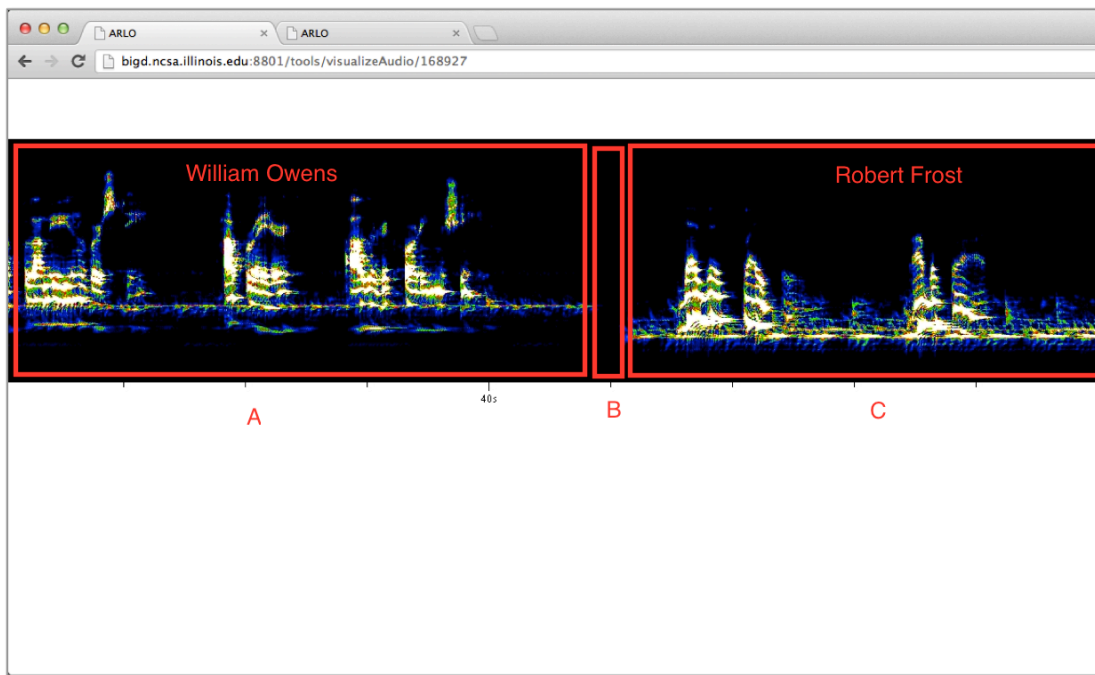**Figure 1.** A "Tagged" Spectrogram in ARLO

After the poets and scholars had used these TEI terms in ARLO to label two-second samples from PennSound's 5497 hours of audio, however, they determined that the list was insufficient for adequately describing the sonic aspects of their audio texts.

[20]

The issue, as described by the participants, was that ARLO's prototype environment for labeling required them to label snippets of sound as though they were static objects while they understood the audio texts to be dynamic events that reflected a relational context. For instance, one participant might mark a two-second sample "Beatable" with a "High" pitch and another might classify the same sample as "Arrhythmic" with a "Low" pitch. This seeming discrepancy was the result of differing perspectives on the sounds since "High" is only perceived as "High" in relation to "Low" and "Arrhythmic" is only meaningful in relation to "Beatable."[4] The participants also wanted to label larger contexts rather than snippets and to tag the recording scenario (such as the sound of the room) over time; the perceived gender of the speaker so that they could mark why they thought a pitch was higher or lower (in relation to how a voice perceived as male or female might register); and the genre of a noise (such as musical or lilting voice) as they perceived it. It quickly became clear that the TEI terms did not provide a way for the participants to label their audio texts as situationally contingent or in a way they felt adequately represented the various audio-textual dynamics they could perceive based on their perspectives in a particular moment in time. Consequently, while participants had rationales of audio-text they wished to access and analyze, their experience in this pilot study seemed to indicate that these rationales were computationally intractable or non-computable in the ARLO system.

[21]

4.2 Audio textuality in the TEI

While the TEI's "Transcriptions of Speech" guidelines represent one of the few rich content models for reproducing, transmitting, and displaying the verbal content of audio texts in the digital humanities [Johansson 1995], it was originally developed to produce transcripts or written representations of a stretch of speech. According to extant meetings notes and draft versions, speech encoding guidelines were not originally included in the original TEI P1 guidelines [Sperberg-McQueen and Burnard 1990]. Indeed, when tasked with creating recommendations for its inclusion, the Spoken Text Working Group (STWG) purposefully restricted their guidelines, as much as possible, to the verbal aspects of a recording [Johansson et al. 1991a]. Consequently, they labeled prosodic and paralinguistic features such as the vocal gestures Bernstein describes including "quasi vocal things such as laughter, quasi lexical things such as 'mm'" [Johansson et al. 1991b] as well as "speaker overlap, pauses, hesitations, repetitions, interruptions", uncertainty, and

[22]

context [Johansson 1995] as "problems." Reflecting these early deliberations, the current guidelines still focus on the verbal content of recordings, stating conclusively that "speech regarded as a purely acoustic phenomenon may well require different methods from those outlined here" [TEI Consortium 2014]. To date, these different methods have never been produced within the context of the TEI.

From the extant meeting notes surrounding the development of these guidelines, it appears that the TEI STWG found time-based sound dynamics such as pitch, speed, and tone meaningful, but impractical to represent given the TEI's structure. Notes from the working group's 1991 meeting reflect the group's knowledge of the relationship between para-linguistic content and cultural studies in the works of [Svartvik and Quirk 1980] and [Tedlock 1983], which are heavily cited, but also the understanding that the STWG had to compromise representing these audio-text dynamics in order to stay consistent with the TEI's underlying OHCO model of textuality and its subsequent use of hierarchical XML (EXtensible Markup Language) elements. The STWG call "performative features such as pitch, speed and vocalization...as analogous to rendition in written texts" and suggest that they be marked similarly in TEI-marked transcriptions, "using milestone tags such as `<pitch.change>`, `<speed.change>` etc." [Johansson et al. 1991b]. Likening these performative features to type changes such as italicized or underlined text, the note suggest "milestone" tags that identify these sonic events as clear points in time with a start and an end point instead of as dynamic events that happen in concert with other patterns across time. The end result is similar to the instances in which structural tags produced in the TEI guidelines for text such as `<p>` for paragraph or `<l>` for line cut off themes across novels or truncate rhyming schemes across poems.

First marking these sonic phenomena as moments of "change" and, later, as spans to "highlight" [TEI Consortium 2004], the STWG eventually relegated the descriptors outlined above to attributes within the "shift" element (`<shift>`), where they still appear in the P5 TEI Guidelines today [TEI Consortium 2014]. Accordingly, the encoder is not only encouraged to mark these dynamic sound attributes in the encoded transcript as defined points in time but also as "shifts" *outside* the "normal" speaking mode [Johansson et al. 1991b]. While the guidelines were "not intended to support unmodified every variety of research undertaken upon spoken material now or in the future", these choices show the prominence that models for verbal content in the form of transcriptions have had in the development of content model standards for audio texts in the digital humanities [TEI Consortium 2014]. Like the scholars before them who critiqued the OHCO model of textuality for poetry, our HiPSTAS participants have indicated that these rationales of audio-text cannot adequately represent rationales of audio-text that include the situationally contingent and time-based vocal gestures of poetry performance recordings.

4.3 Audio texts in the Bibliographic Framework Initiative

The Bibliographic Framework Initiative (BIBFRAME) was founded by the Library of Congress (LC) as a movement to replace the LC's MARC (Machine-Readable Cataloging) standard[5] and represents an example of current thinking concerning how cultural artifacts such as texts and audio-texts are described for discovery in large repositories. Like the FRBR model, the BIBFRAME model seeks to establish a standard model for standards in "all aspects of bibliographic description, data creation, and data exchange . . . includ[ing] accommodating different content models and cataloging rules, exploring new methods of data entry, and evaluating current exchange protocols" [Bibliographic Framework Initiative]. Like FRBR, the BIBFRAME model also has a central imperative to create models that represent and maximize a user's ability to find, identify, select, and obtain library holdings such as books and audio files by "differentiat[ing] the conceptual work (a title and author) from the physical details about that work's manifestation (page numbers, whether it has illustrations)" [Bibliographic Framework Initiative].

The aspect of the BIBFRAME initiative of concern for this piece is a recent report created by members of the initiative called "BIBFRAME AV Modeling Study: Defining a Flexible Model for Description of Audiovisual Resources" [Van Malssen 2014], which is markedly different than FRBR. While the author of the "BIBFRAME AV Modeling Study" is, like FRBR's authors, very keen to incorporate alternate models that meet both public and scholarly needs for access and discovery methods, her primary model of reference, unlike FRBR's or the TEI's, is time-based audiovisual resources rather than verbal texts.

23

24

25

26

As the report indicates, AV resources have medium-specific characteristics that must be taken into consideration in any digital representation. First, they are by nature event-centric and carrier dependent. That is, what is central to the rationale of such recordings is that an event or action (or a continuum of events and actions) that took place in a time and at a place was recorded or "fixed" to a carrier such as a film reel or an audio tape. Van Malssen notes that "while machine dependency is not unique to time-based media, it is an unextractable attribute" that must be taken into account when considering provenance but also for access and discovery in terms of creating playback opportunities at the right speed or color resolution [Van Malssen 2014, 6]. Second, AV resources are more often than not created by and contributed to by many collaborators. Third, these materials are often unique in that they have never been published [Van Malssen 2014, 7]. Fourth, AV resources can be singular items but they are also manifest in the aggregate as collections and can appear on multiple carriers. For example, albums have multiple tracks; songs appear in multiple collections; and long interviews or films are often recorded across multiple tapes or reels [Van Malssen 2014, 8]. Fifth, the audio and moving image media we see today have complicated relationships to their original recordings. Most have been migrated across media formats as older media degenerates and new media evolves. Historical recordings migrated to reel-to-reel tapes, for example, may have first been recorded on glass discs that were shorter in length. Thus, the reel-to-reel tapes do not have a one-to-one relationship to the carriers that originally held the recording. Or, film reels may have been conserved or enhanced in later generations after decades of poor quality viewings. Because of these important characteristics of AV materials, Van Malssen argues that models for cataloguing (and thus discovering) AV must reflect a primary concern with modes of creation and access rather than simply a focus on particular types of content [Van Malssen 2014, 4].

Due to these issues, Van Malssen recommends an event-centric content model for AV resources that takes the physically situated context (e.g., the recording and playback circumstances) into account. If expressions can be described as either work- or event-centric, it would allow for "the inclusion of works, or some work elements as part of the event, but does not require a work be present" [Van Malssen 2014, 24]. This distinction between *event* and *work* is exemplified by the difference between an audio recording of a battle that took place in World War II and a filmed reenactment of the same battle. A description of the battle recording would be *event-centric* since the event took place in space and time and is the object of description at the content level; the filmed re-enactment would be more like the traditional notion of the work in the mind of the creator [Van Malssen 2014, 18]. Van Malssen notes that not treating these catalog descriptions differently results in a blurring of resources: both event and the reenactment would employ the same subject terms (World War II) even though the first *is* the event while the second uses the event as content.

For audio texts such as recordings of poetry performances, an event-centric description could reshape processes for discovery and possibilities for research. For instance, poetry readings that are recitations of a written text could be marked differently than poetry performances. FRBR identifies an *expression* as a different edition or translation of a work and the *manifestation* as the physical embodiment of that expression with the *item* as "a single exemplar of a manifestation." A marked difference between work and event would be salient in poetry performance studies where the performance of a poem that is also written could be very similar to or very different from the written text – either a new expression or a whole new work (that is the event). For example, in Ken Sherwood's analysis of the recordings and transcriptions of Amiri Baraka, Kamau Brathwaite, and Cecilia Vicuña, he speaks of "emergent performance events" as different works in comparison to poetry recitations, which he considers oral versions (or variant expressions) of the written text [Sherwood 2006]. A performance, Bernstein argues "opens up the potential for shifting frames, and the shift of frame is itself perceived as a performative gesture . . . the implied or possible performance becomes a ghost of the textual composition" [Bernstein 2011, 127]. A model that could be either event- or work-centric could serve to help identify these kinds of hybrid situations in which an event gives rise to a work or vice versa, creating an entirely new, but related, entity.

It is clear that our print-based, logocentric biases preclude us from facilitating new kinds of discovery and analysis with alternate media, but seeing these biases also makes it apparent that we are building models of textuality in general that remain insufficiently poor. The attempt in the HiPSTAS project to create an infrastructure for analyzing audio patterns in audio texts illustrates the importance in the humanities for rationales of audio text that engage audio as events that may make meaning differently by means of markedly different reception scenarios. Finding and making meaning with audio

texts can include a consideration for sound dynamics that can only be perceived through time such as pitch, timbre, and intonation as well as the importance of material recording scenario artifacts, all of which could be considered "noise" and essentially rendered undiscoverable in the context of systems that model all texts as if they are static, verbal documents. As a culture, we have become adept at modeling textuality in terms of a print-based paradigm in which texts are discrete entities (in terms of both time and space) of verbal content. We are arguably less adept at modeling time-based or three-dimensional media, which we still insist on fixing in time and translating into words, either through metadata or through transcripts.[6] Discussions concerned with access and scholarship with AV materials provides for an opportunity for expanding our thinking about these "other" texts but also how representations of traditional (or book-like) verbal texts are equally insufficient in the digital environment.

5. Towards [Other] Rationale[s] of [Audio] Text

The expanding web, advancing opportunities for networking within and across library collections through Linked Data, and our increasingly complex needs for discovering and analyzing all types of library holdings has resulted in a general re-evaluation of how we model all types of cultural artifacts through information systems. Situating the above reflections on the rationale of audio text in the context of Renear and Durbin's recommendations for roles in FRBR and the AV model recommendations for describing events in the LC's BIBFRAME standard provides for an opportunity to reconsider methods for representing textuality more generally as always situated and contingent in information systems. Specifically, the following section introduces three aspects of textuality that could be foregrounded through event-centric content and discovery models: collective intentionality, emergent performativity, and indexical performativity.

31

5.1 Collective intentionality

With an event-centric model, textual events could be described as social objects that reflect collective intentionalities. Arguing that the role a "natural" object plays is determined by contingent social circumstances, Renear and Durbin cite J.R. Searle's understanding of these circumstances as the "collective intentionality" of producers, transmitters, and consumers who shape the role an object plays in a system ([Searle 1995], quoted in [Renear and Durbin 2007]). This notion of collective intentionality is also essential in social text theories such as Martha Nell Smith's ideas concerning triangular textuality ("the influence of biography, reception, and textual reproduction") in Emily Dickinson's work [Smith 1992, 2]; in the context of the Ivanhoe Game in which the reader's interpretations are essential to the perceived rationale of text [McGann 2001] [Drucker and Rockwell 2003]; or in D.F. McKenzie's insistence that texts serve as signposts for spatio-temporality, as they alert "us to the roles of institutions and their complex structures in affecting the forms of social discourse, past and present" [McKenzie 1999, 15]. In defining each manifestation as the fixing or recording of socio-textual events, McGann, McKenzie, Smith, and others are discussing texts as social objects in much the same way as Renear and Durbin who are arguing that the social and cultural circumstances that help produce texts and our understandings of a text should be included in computational models for texts.

32

While it would seem that such social circumstances are beyond the codifications of a bibliographic standard for access, there are emerging models for discovering AV that are promising. Van Malssen cites *indecs*, a project of the European Community Info 2000 initiative, as an AV model that adequately "places emphasis on the role that events play in the creation of a resource" by using the expression entity to describe events as "'creating' events (an event which results in the making of a creation), 'using' events (events which results [sic] in the use of a resource), or 'transforming' events which involve the use of one creation in the making of another" [Van Malssen 2014, 19]. Marking the role of event is markedly different even than the promising W3C Provenance Data Model (PROV-DM), which is a framework that includes "information about entities, activities, and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability or trustworthiness" [PROV-DM]. Designed to model the relationships involved in the creation of items, PROV-DM is organized into six components that include the following: "(1) entities and activities, and the time at which they were created, used, or ended; (2) derivations of entities from entities; (3) agents bearing responsibility for entities that were generated and activities that happened; (4) a notion of bundle, a mechanism to support provenance of provenance; (5) properties to link entities that refer to the same thing; and, (6) collections forming a logical structure for its members" [PROV-DM]. Like the FRBR model, however, PROV-DM does not designate the role of events in terms such as "creating events," "transforming events," and "using events" but

33

rather the manifestations such as texts, paintings, or photographs that result from these events. Thus, while a more robust model such as PROV-DM can provide a model for text that foregrounds the results of collective intentionalities (and thus records its occurrence), the event-centric information concerning the roles these events play in relationship to each other and to the objects they effect remains obfuscated.

5.2 Emergence

A model of textuality that represents text as a spatial and temporal phenomenon might allow for interactions and representations in a digital environment that, rather than insisting on fixity, foreground principles of emergence. For example, the AV model recommendations for BIBFRAME note that FRBR does not provide a model for describing an aggregate manifestation that comprises two or more distinct expressions such as when two different parts of a recording are recorded or "fixed" at different times and different places or when different tracks on a recording are compilations of wholly different works, a scenario that is not far removed from similar instances of volumes of poetry or short stories, which have been republished and reissued. This inadequacy comes into greater relief as we begin to think about the creating, using, and transforming events that give rise to whole new works.

The recommendations for a BIBRAME AV model attempts to expand on the notion of fixity that is inherent to the FRBR model. Discussing a digital video file that might have singular or multiple video and audio tracks plus a subtitle track, the AV model recommendations cite the concept of *essence* as it is used in the PBCore (Public Broadcasting Metadata Dictionary Project) and EBU (European Broadcasting Unit) Core models for broadcast collections as a more robust concept for an event-centric model since it typically includes information about both the content and carrier as well as enabling the description of multiple instantiations of a given content item, such as a broadcast that is repeatedly aired at different times. The AV model recommendations favor the essence model as one that "is as much about the bitstream or signal as it is about the content carried in that stream" because it "is expressed in these models through sets of carrier sub-elements . . . which can be repeated for different essence types found in that carrier" [Van Malssen 2014, 18].

This notion of essence corresponds to John Bryant's definition of verbal texts as "fluid" or as a "flow of energy" rather than a product or a "conceptual thing or actual set of things or even discrete events" [Bryant 2002, 61]. According to this definition, the aggregated text comprises multiple versions in manuscript and print, various notes and letters and comments of contemporaries or current readers, as well as an element of what I have called *textual performance* [Clement 2011] or a concept of textuality as it is manifested or performed in real time and space with a collaborative audience. By developing models that expand notions of fixity as an essential element, we allow manifestations to be aggregate and contingent, a representation for text that allows for new modes of discovery and access.

5.4 Indexical performativity

As discussed, according to the recommendations for a BIBFRAME AV model, an AV object should always be described in terms of the event of its realization, in terms of how it has been and could be realized on a particular carrier or format (such as a vinyl record or a CD), or in terms of a particular playback system (such as a record player or CD player). As such, every manifestation becomes indexical because each points to these events or technological protocols that generated the item. As such, the nature of an object's identification as an "item" would be based in part on the information surrounding the contingent circumstances of its realization. For instance, in order to enable "authentic access and preservation," the AV model recommendations cite a need for software systems that not only describe granular features of certain formats such as "audio reel, number of tracks, playback speed, reel size, etc.", but also the technical characteristics of carriers in order "to enable identification of the appropriate playback equipment" that can respond to or make manifest these features [Van Malssen 2014, 24].

Accordingly, these aspects of an object's indexicality are also performative in a digital environment since they create how we understand the structure of the event of an object's instantiation. Wolfgang Ernst describes this kind of performativity in the digital environment as the extent to which code plays a role as text, as image, as audio, and so on; from this perspective, "rigid text is replaced by an operative mathematics" [Enrst 2012, 87]. By drawing attention to the

manner in which code effects commensurability among digital "items" such as text, images, and sound, Ernst considers "the structure of an archive whose essence, the closer one looks [as] less the archived material per se than a dynamic conception of the idea of the archive" [Enrst 2012, 83]. This is to say that the very nature of digital materiality requires that a model for "item" (through attribute and values) be applied on the level of the bitstream in order for any text to be realized as such. This aspect of the item's indexical performativity is also identified in Renear and Durbin's assertion that "items are the things, whatever their nature (physical, abstract, or metaphorical), which play the role in bibliographic control that FRBR assigns to items" [Floyd and Renear 2007]. In both cases, the attributes and relations of an item (whether on the level of its physical or digital instantiation) that effect its manifestation are also what would define its role as an item in a system like FRBR or BIBFRAME and our understandings of its "itemness".

39

When modeled as event-centric, these attributes and relations of an item are directly affected by changing data models that demand updated software and hardware with changing protocols and algorithms for the constant realizing of items over time and across systems. For instance, as Jonathon Sterne argues, "the transformations effected by MP3 encoding are themselves heavily-directed cultural practices. MP3s contain within them a whole philosophy of audition [based on the limitations of human hearing] and a praxeology of listening . . . [that] emphasizes distraction over attention and exchange over use" [Sterne 2006, 828]. Noting that the MP3 is a container for an audio digital recording that has been compressed with the use of a mathematical model that takes human auditory perception into account, Sterne writes that the "most compelling part of the MP3 is the psychoacoustic model encoded within it . . . It preemptively discards data in the sound file that it anticipates the body will discard later, resulting in a smaller file" [Sterne 2006, 833]. Depending on what aspect of the sound recording one seeks to address, therefore, the MP3, which is typically categorized in traditional models such as FRBR as a derivative item of the same manifestation, could be understood as actually realizing a different event: one without the soundscape that exists beyond human hearing.

40

This notion of an event-centric indexicality can be made more broadly applicable to all kinds of digital textuality. Michael Witmore has written that "a text is a text because it is massively addressable at different levels of scale . . . one can query a position within the text at a certain level of abstraction" such as a word or a line, but also on the level of a theme or a character [Witmore 2010]. In a digital context in which we are attempting to model textuality for discoverability and analysis, a text can also be understood as massively and differently addressable and therefore indexical according to how the protocols we design in the system instantiate it at a given point and time. Consider Witmore's definition of text addressability in relation to Sterne's description of the MP3 and McGann's "Rationale of Hypertext" in which a central identifying feature of the hypertext is that "[un]like a traditional book or set of books, the HyperText need never be 'complete'"; rather, the hypertext by nature "will evolve and change over time, it will gather new bodies of material, its organizational substructures will get modified"; and, "[u]nlike a traditional edition, a HyperText is not organized to focus attention on one particular text or set of texts. It is ordered to disperse attention as broadly as possible" [McGann 2001, 71]. With such a text, Witmore's levels of abstraction are exponentially complicated by the digital text's complexities. The digital text can become differently discernible and interpreted by different audiences with each different manifestation or code performance. In web pages created "on the fly" with PHP (Hypertext Preprocessor), for example, the same URI (Uniform Resource Identifier) or URL (Uniform Resource Locator) could take a reader to a different textual instantiation each time. As McGann reminds us, with hypertext, "[o]ne is encouraged not so much to find as to make order -- and then to make it again and again, as established orderings expose their limits . . . [the edition] will incorporate and then go beyond its initial design and management" [McGann 2001, 71]. Event-centric attributes and relations have an indexical performativity across all texts (written, verbal and AV materials) in that the technological context in which the text has been created and "served" or delivered to the scholar is not only reflected in the item, it creates our understanding of the item's *itemness* (such as "a page created on the fly") and therefore directly impacts our understandings of its textuality more generally.

6. Conclusion

There is an aside in McGann's "Rationale of Hypertext" that almost goes unnoticed. In discussing the myriad extensibilities that hypertext employs that "go beyond its initial design," McGann writes, "*Someone will have to manage it*..." [McGann 2001, 71]. This conclusion concerns that "someone" or, rather, the realities and practicalities of managing computationally tractable as well as authentic and recognizable, texts in an information system. Renear and Durbin, for

41

example, note that FRBR is a guideline for practice, an "extraordinarily promising and compelling...general ontology," which never claims its entities *are* types — and thus, Renear and Durbin concede, their argument for articulating them as *roles* could be considered, at best, an academic exercise, and, at worst "willfully obtuse" [Renear and Durbin 2007]. Similarly, Van Malssen writes that the recommended BIBFAME AV model is meant to be "intuitive enough so a trained professional user is able to interpret it and easily make decisions about how to apply it to a given content type" [Van Malssen 2014, 25]. The recommendations are an attempt to suggest a model that is flexible enough to describe film and audio as well as books in an effort to create practical guidelines or guidelines for real information infrastructures that are managed by and used by real practitioners and scholars.

Yet, a concern for the ontological issues that shape information management systems is increasingly important since widely used ontologies like FRBR and BIBFRAME and semantics like the TEI often impact other practicalities such as data interchange in Linked Data and the Semantic Web. For instance, projects such as ARC (Advanced Research Consortium), which is the umbrella organization for NINES, 18thCONNECT, ModNets and other communities who are trying to both shape and facilitate digital scholarship in the humanities,[7] model their RDF (Resource Description Framework) data models for Linked Data on the Dublin Core standards, which are, in turn, heavily influenced by FRBR [Chaudhri 2009].

42

As such, this discussion serves as a provocation rather than a practical how-to. I am noting here that digital humanities has a tradition of problematizing our understanding of textuality through the design of information systems. I am also suggesting that these efforts have been almost exclusively performed in the context of rationales for verbal texts that still, in many (if not most) cases, look like books or pages of books. Scholars within this tradition have an opportunity to ask more questions and to create different, more encompassing solutions within the context of designing systems for AV materials.

43

To conclude, both the Council for Library and Information Resources (CLIR) and the Library of Congress (LC) have issued reports detailing the dire state of access and preservation with sound recordings. Both have called for "new technologies for audio capture and automatic metadata extraction" [Rumsey et al. 2004] with a "focus on developing, testing, and enhancing science-based approaches to all areas that affect audio preservation" [The Library of Congress 2012, 15] in order to help relieve backlogs of undescribed (even though digitized) audio collections and to facilitate better means for access and discovery. One of the results of these reports is an urgency to re-evaluate how we develop metadata frameworks that facilitate access to audio. One means of re-evaluation is a reconsideration of our rationales of text. What if all entities were modeled as dynamic events? Would notions of collective intentionality, emergence, and indexical performativity foreground our searches and shape what we are able to find? Text is central to the Humanities, but as the editor of this special issue notes, "digitization remediates all analog sources into a common binary format," a situation which serves as an invitation to think through these remediations towards the development of future information systems that will continue to shape new and different rationales of text.

44

Acknowledgments

45

Notes

[1] Audio is best defined as sound that has been recorded, transmitted, or reproduced.

[2] A computational model of text, in Willard McCarty's terms, would be both completely explicit and absolutely consistent as well as manipulable.

[3] The instantaneous energy is then factored by summing the fork's potential energy or the deflection of the fork and its kinetic energy based on the speed of the movement, per second.

[4] These results are discussed in detail in [Clement 2014].

[5] The MARC standard, which was originated by the LC, is a set of digital formats for describing and cataloguing items that has been an international standard since the 1970s.

[6] See more discussions of these perceived inadequacies in [Good 2006] and [Nyhan and Flynn 2014].

[7] ARC is described at http://idhmc.tamu.edu/arcgrant/.

Works Cited

Bauman 1975 Bauman, R. "Verbal Art as Performance." In *American Anthropologist*, New Series, 77, no. 2 (June 1975): 290-311.

Bernstein 2011 Bernstein, C. *Attack of the Difficult Poems: Essays and Inventions*. University Of Chicago Press, 2011.

Bibliographic Framework Initiative Bibliographic Framework Initiative. Library of Congress (May 15, 2014). http://www.loc.gov/bibframe/.

Boruszak 2014 Boruszak, J. https://sites.google.com/site/nehhipstas/project-pages/jeff-boruszak

Briet 2006 Briet, S. *What is Documentation? English Translation of the Classic French Text*. Translated and edited by R. E. Day and L. Martinet. Lanham, MD: Scarecrow Press, 2006.

Bryant 2002 Bryant, J. *The Fluid Text: A Theory of Revision and Editing for Book and Screen*. Ann Arbor: University of Michigan Press, 2002.

Bryant 2011 Bryant, J. "Where Is the Text of America? Witnessing Revision and the Online Critical Archive." In *The American Literature Scholar in the Digital Age*, edited by Amy E. Earhart and Andrew Jewell. Ann Arbor: University of Michigan Press, 2011.

Buckland 1997 Buckland, M. "What is a 'document'?" *Journal of the American Society for Information Science* 48, no. 9 (1997): 804-809.

Buzzetti and McGann 2006 Buzzetti, D. and McGann, J. J. "Critical Editing in a Digital Horizon." In *Electronic Textual Editing,* edited by Lou Burnard, Katherine O'Brien O'Keeffe, and John Unsworth. New York: Modern Language Association of America, 2006. http://www.tei-c.org/About/Archive_new/ETE/index.xml.

Chaudhri 2009 Chaudhri, Talat. "Assessing FRBR in Dublin Core Application Profiles." Ariadne 58 (2009): n. pag. http://www.ariadne.ac.uk.

Clement Clement, T. E. "When Texts of Study are Audio Files: Digital Tools for Sound Studies in DH" In *A New Companion to Digital Humanities* (Blackwell Companions to Literature and Culture). Susan Schreibman, Ray Siemens and John Unsworth (eds.) (accepted for publication).

Clement 2011 Clement, T. E. "Knowledge Representation and Digital Scholarly Editions in Theory and Practice." *Journal of the Text Encoding Initiative* 1, no. 1 (June 2011). http://jtei.revues.org/203.

Clement 2014 Clement, T. E. "The Ear and the Shunting Yard: Meaning Making as Resonance in Early Information Theory." *Information & Culture* 49.4 (2014): 401-426.

Clement et al. 2014 Clement, T., Tcheng, D. Auvil, L. and Borries, T. "High Performance Sound Technologies for Access and Scholarship (HiPSTAS) in the Digital Humanities" *Proceedings of the 77th Annual ASIST Conference*, Seattle, WA, October 31-November 5, 2014.1.

Council on Library and Information Resources and The Library of Congress 2012 Council on Library and Information Resources and The Library of Congress. *National Recording Preservation Plan*. Washington, DC: Council on Library and Information Resources and The Library of Congress, 2012.

DeRose et al. 1990 DeRose, S. Durand, D. G., Renear, A. H. "What Is Text, Really?" *Journal of Computing in Higher Education* 1, no. 2 (1990): 3–26.

Drucker 2002 Drucker, J. "Theory as Praxis: The Poetics of Electronic Textuality." *Modernism/modernity* 9, no. 4 (2002): 683–91. doi:10.1353/mod.2002.0069.

Drucker and Rockwell 2003 "Introduction; Reflections on the Ivanhoe Game." *Text Technology* 12.2 (2003): vii-xviii.

Enrst 2012 Ernst, W. *Digital Memory and the Archive*. Minneapolis: Univ Of Minnesota Press, 2012.

Enstrom 1993 Enstrom, D. A. "Female Choice for Age-Specific Plumage in the Orchard Oriole: Implications for Delayed Plumage Maturation." *Animal Behaviour* 45, no. 3 (March 1993): 435–42. doi:10.1006/anbe.1993.1055.

Feinberg 2010 Feinberg, M. "Two kinds of evidence: how information systems form rhetorical arguments." *Journal of Documentation* 66, no. 4 (2010): 491-512.

Floyd and Renear 2007 Floyd, I. and Renear, A. H. "What Exactly is an Item in the Digital World?" Poster presented at the Annual Meeting of the Association for Information Science and Technology, Milwaukee, Wisconsin, October 19-24, 2007.

Frohmann 2004 Frohmann, B. *Deflating Information: From Science Studies to Documentation*. Toronto: University of Toronto Press, Scholarly Publishing Division, 2004.

Goldfarb 1981 Goldfarb, C.. "A generalized approach to document markup." Proceedings of the ACM SIGPLAN--SIGOA Symposium on Text Manipulation. New York: ACM (1981): 68-73.

Good 2006 Good, F. "Voice, ear and text: words, meaning and transcription." In R. Perks & A. Thomson, eds. *The Oral History Reader*. New York: Routledge, 2006: 362–373.

Johansson 1995 Johansson, S. "The Encoding of Spoken Texts." *Computers and the Humanities* 29, no. 2 (1995): 149–58.

Johansson et al. 1991a Johansson, S. Burnard, L., Edwards, J., Rosta, A. "TEI AI2 W1 Working paper on spoken texts University College London." TEI Consortium (October, 1991a). http://www.tei-c.org/Vault/AI/ai2w01.txt.

Johansson et al. 1991b Johansson, S. Burnard, L., Edwards, J., Rosta, A. "TEI AI2 M1 Minutes of Meeting Held at University of Oslo." TEI Consortium (9-10 August 1991b). http://www.tei-c.org/Vault/AI/ai2m01.txt.

Kirschenbaum 2008 Kirschenbaum, M. G. *Mechanisms: New Media and the Forensic Imagination.* Cambridge, MA: MIT Press, 2008.

Liu 2004 Liu, A. "Transcendental Data: Toward a Cultural History and Aesthetics of the New Encoded Discourse." *Critical Inquiry* 31, no. 1 (2004): 49–84.

McCarty 2004 McCarty, W. "Modeling: A Study in Words and Meanings." In *Companion to Digital Humanities* (Blackwell Companions to Literature and Culture), edited by Susan Schreibman, Ray Siemens, and John Unsworth. Oxford: Blackwell Publishing Professional, 2004. http://www.digitalhumanities.org/companion/.

McGann 1983 McGann, J. J. *A Critique of Modern Textual Criticism*. Chicago: University of Chicago Press, 1983.

McGann 1991 McGann, J. J. *The Textual Condition*. Princeton, N.J: Princeton University Press, 1991.

McGann 1995 McGann, J. J. "The Rationale of HyperText," May 6, 1995. http://www.iath.virginia.edu/public/jjm2f/rationale.html.

McGann 2001 McGann, J. J. *Radiant Textuality: Literature After the World Wide Web*. New York: Palgrave, 2001.

McKenzie 1999 McKenzie, D. F. *Bibliography and the Sociology of Texts*. Cambridge University Press, 1999.

Mustazza 2014 Mustazza, C. https://sites.google.com/site/nehhipstas/project-pages/chris-mustazza.

Nyhan and Flynn 2014 Nyhan, J. and Flynn, A. "Oral History, audio-visual materials and Digital Humanities: a new 'grand challenge'". In AV in DH Workshop Proceedings, Digital Humanities Conference. Lausanne, Switzerland, July 2014. https://avindh2014.wordpress.com/abstracts/#ab4

Otlet 1990 Otlet, P. *International Organisation and Dissemination of Knowledge: Selected Essays of Paul Otlet*. Amsterdam: Elsevier for the International Federation of Documentation, 1990.

PROV-DM "PROV-DM: The PROV Data Model", Luc Moreau, Paolo Missier (eds.), W3C Recommendation, 30 April 2013, http://www.w3.org/TR/prov-dm/ Latest version available.

Renear 1996 Renear, A. H., Mylonas, E., and Durand, D. "Refining Our Notion of What Text Really Is: The Problem of Overlapping Hierarchies." In *Research in Humanities Computing*, edited by Nancy Ide and Susan Hockey. Oxford University Press, 1996. http://hdl.handle.net/2142/9407.

Renear 2004 Renear, A. H. "Text Encoding." In *A Companion to Digital Humanities*, edited by Susan Schreibman, Ray Siemens, and John Unsworth. Blackwell Publishing Ltd, 2004, 218–39. http://onlinelibrary.wiley.com.ezproxy.lib.utexas.edu/doi/10.1002/9780470999875.ch17/summary.

Renear 2005 Renear, A. H. "Text from several different perspectives, the role of context in markup semantics." In *Atti della conferenza internazionale CLiP 2003, Computer Literacy and Philology (Firenze, 4–5 December 2003)*, edited by C. Nicolas and M. Moneglia. Florence: University of Florence Press, 2005.

Renear and Durbin 2007 Renear, A. H. and Durbin, D. "Three of the Four FRBR Group 1 Entity Types are Roles, not Types." In *Proceedings of the American Society for Information Science and Technology* 44, no. 1 (2007): 1-19.

Rettberg 2014 Rettberg, E. https://sites.google.com/site/nehhipstas/project-pages/eric-j-rettberg.

Rumsey et al. 2004 Rumsey, A. S., Allen, D. R., and Allen, K. Council on Library and Information Resources. *Survey of the State of Audio Collections in Academic Libraries*. Washington, D.C.: Council on Library and Information Resources, 2004. http://catalog.hathitrust.org/Record/005405973.

Searle 1995 Searle, J. R. *The Construction of Social Reality*. New York: The Free Press, 1995.

Sherwood 2006 Sherwood, K. "Elaborate Versionings: Characteristics of Emergent Performance in Three Print/Oral/Aural Poets." *Oral Tradition* 21, no. 1 (2006): 119-147.

Smith 1992 Smith, M. N. *Rowing in Eden: Rereading Emily Dickinson*. Austin: University of Texas Press, 1992.

Sperberg-McQueen and Burnard 1990 Sperberg-McQueen, M. and Burnard, L. eds. *Guidelines for the Encoding and Interchange of Machine-readable Texts*. Draft version 1.0. Chicago and Oxford: Association for Computers and the Humanities/Association Computational Linguistics/ Association for Literary and Linguistic Computing, 1990.

Sterne 2006 Sterne, J. "The MP3 as Cultural Artifact." *New Media and Society* 8, no. 5 (November 2006): 825-842.

Sterne 2012 Sterne, J. "Sonic Imaginations." In Sterne, J. (ed.), *The Sound Studies Reader*, edited by Jonathon Sterne. New York: Routledge, 2012: 1-18.

Svartvik and Quirk 1980 Svartvik, J. and Quirk, R. eds. A Corpus of English Conversation. *Lund Studies in English* 56. Lund: Lund University Press, 1980.

TEI Consortium 2004 TEI Consortium, eds. *The XML Version of the TEI Guidelines: Notes for TEI P4 Guidelines for Electronic Text Encoding and Interchange* XML-compatible edition. Text Encoding Initiative (June 2004). http://www.tei-c.org/Vault/P5/1.0.1/doc/tei-p4-doc/html/index-notes.html.

TEI Consortium 2014 TEI Consortium, eds. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Version 2.6.0. TEI Consortium (2014). http://www.tei-c.org/Guidelines/P5/.

Tedlock 1983 Tedlock, D. *The Spoken Word and the Work of Interpretation*. University of Pennsylvania Publications in Conduct and Communication. Philadelphia: University of Pennsylvania Press, 1983.

The Library of Congress 2012 The Library of Congress. "Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services." Washington, DC: The Library of Congress, 2012.

Tillet 2004 Tillet, Barbara "FRBR: A Conceptual Model for the Bibliographic Universe." Library of Congress Cataloging Distribution Service, 2004.

Van Malssen 2014 Van Malssen, K. "BIBFRAME AV Modeling Study: Defining a Flexible Model for Description of Audiovisual Resources." Library of Congress (May 15, 2014). http://www.loc.gov/bibframe/.

Witmore 2010 Witmore, M. "Text: A Massively Addressable Object." *Wine Dark Sea* (December 31, 2010). http://winedarksea.org/?p=926.