Yet of Books There Are A Plenty: The Bibliography of Literary Data

Nigel Lepianka

May 16, 2017

Chapter 1

The Rationale of Aggregation

An enumerative bibliography is an interpretation of scale. Bibliographies are not the sole genre to attempt to do so; their nature as an interpretation is that they leave open gaps or alternative ways for the immensity of information to be comprehended.[1] A list of texts or works, whether multi-volume monographs or the appendage of an article or smaller work, exist on the one hand to show the labor both required and necessary for research. On the other hand, it reveals to the reader of the work what does exist, and can be used at the reader's discretion to further expand or complicate the discussion at hand. In particular, bibliographies are an attempt to recognize the vastness of text, primarily published text. Every bibliography is an admission of the enormity of literature, regardless of how limited or broad the definition of that term. For example, Lyle H. Wright's *American Fiction* bibliography is far more constrained in its scope compared to Jacob Blanck's *Bibliography of American Literature.* Wright's work however, still manages to fill three volumes spanning from the years 1774 to 1900 and listing more than 12,000 titles.[2] Despite, and even because of, this attempt to grapple with the enormity of a subject such as the production of early American literature, errors, omissions, and erroneous inclusions into Wright's bibliography exist, both consciously and unconsciously.

What Wright, Blanck, and other bibliographers attempt to do in their bibliographies can not be said to simply list books they find, but, rather to organize them into a schema that is understood by the bibliographer to best aid the reader of the list. For what purpose the reader may be viewing

1. Other examples I would argue as print-based interpretations of scale would include concordances, which attempt to form a comprehensive reference of key terms and their usage in writing, or sometimes whole vocabularies, for a particular category of writing. Anthologies represent an attempt at trying to assemble representative works of a given category often as teaching aids. Dictionaries and thesauruses are attempts to formalize and fit into a particular paradigm the denotative properties of a language's words. Indexes as well attempt to infer what a reader will want to look up or find within a text, and so catalogs a list of assumed key terms. All of these particular genres of reference work in some way assume the task of making a large amount of information more digestible to readers.

2. For comparison, Blanck's *BAL* fills nine volumes and lists more than 23,000 entries. The descriptions Blanck provides for each entry are also more substantial, whereas Wright, as will be discussed later, is more discriminate in what information he includes.

the bibliography is subjective and susceptible to change as scholars pursue different questions and conversations, but must nonetheless be considered in the creation of the bibliography. Bibliographies are assemblages; they are texts like any other that are reliant upon a network of persons each with their own intervention in the process of composition, circulation, and transmission. I use the term assemblage because it concisely points to the wider social nature behind the construction of complex aggregations of data such as bibliographies. Elizabeth Maddock Dillon has employed the term *assemblage* in a literary context to describe the state of agency amongst colonized subjects amidst the Haitian Revolution, remarking on the fact that lack of sovereignty did not necessarily mean a lack of agency.[3] Ryan Cordell has borrowed Dillon's interpretation of the concept and applied it to textual criticism through his Viral Texts Project.[4] For Cordell, an *assemblage* is a text which is "defined by circulation and mutability," and is not beholden to its author, or even its editors and publishers, but can be liberated, reimagined and reinterpreted, by its readers.[5] In this context, Cordell admits, it is an extension of D. F. McKenzie's "sociology of the text", but with a focus on the life of the text after publication rather than before.[6]

My use of *assemblage* is perhaps counter to Cordell's more progressive use in its application to textual criticism and bibliography, but that is because of the properties of a bibliography's composition and publication that influence not just the way readers interact with a text, but in fact, whether or not readers have access to the text at all. A bibliography, as an assemblage is not just the work of the bibliographer who compiled the lists. Certainly, a figure such as a Wright put great effort into compiling three volumes of American fiction, but he was not the sole actor. Wright traveled to several institutions in the process of creating the *American Fiction* bibliography; a total of thirty-one different institutional names can be found in each of the prefatory listing of libraries across the bibliography's three volumes. His descriptions not only reveal this fact, but in fact rely upon it as a testament to the truth of his entries. Each title's description includes credit to the

3. The term is rooted in discussions by Gilles Deleuze, Felix Guattari, Bruno Latour, and "Actor Network Theory", a means of discussing the constructed nature of scientific classification. See Elizabeth Maddock Dillon, "Obi, Assemblage, Enchantment," *J19: The Journal of Nineteenth-Century Americanists* 1, no. 1 (April 22, 2013). 172–178, ISSN: 2166-7438, accessed April 22, 2017, doi:10.1353/jnc.2013.0020, https://muse.jhu.edu/article/504853; Gilles Deleuze and Félix Guattari, *A thousand plateaus : capitalism and schizophrenia. Gilles Deleuze, Félix Guattari ; translation and foreword by Brian Massumi* (Minneapolis : University of Minnesota Press, [1987], 1987), ISBN: 978-0-8166-1401-1; Bruno Latour, *Reassembling the social : an introduction to actor-network-theory. Bruno Latour*, Clarendon lectures in management studies (Oxford ; New York : Oxford University Press, 2005., 2005), ISBN: 978-0-19-925604-4

4. Ryan Cordell, "Reprinting, Circulation, and the Network Author in Antebellum Newspapers," *American Literary History* 27, no. 3 (September 1, 2015). 417–445, ISSN: 0896-7148, accessed April 22, 2017, doi:10.1093/alh/ajv028, https://academic.oup.com/alh/article/27/3/417/85989/Reprinting-Circulation-and-the-Network-Author-in.

5. Ryan Cordell, "Two (of Three) Ways of Looking at C19 Newspaper Exchange Networks," Ryan Cordell, June 22, 2016, accessed April 22, 2017, http://ryancordell.org/research/two-of-three/.

6. At the "What is Critical Bibliography?" panel at MLA 2018 in Philadelphia, MA, the term was discussed, though its meanings, the panel concluded, were variable.

library location(s) in which the title was found. This is an important detail to this, and any, bibliography's composition to make evident. As a resource for scholars that would claim boldly to represent an aggregation of American fiction, it is the responsibility of the bibliographer to locate texts in multiple locations. The aggregation of data in general is itself an act that is defined by its multiplicity. Data that is compiled all from a single source loses most of its justification for having been collected. At best, a bibliography composed of titles from a single library would be a poorly selective catalog. At worst, it would be unreliable in its claims of offering accurate descriptions of the texts listed due to its inherent blindness to alternatives. Thus, when I use the term assemblage, I mean to suggest in the case of texts whose primary function is to serve as an aggregation, that the actors whose hands intervene in the creation of such texts play a role not simply just in the creation of a physical text, but in fact in the use and reception of that text, and, in turn, affecting the ways in which the objects listed are accessed and received.[7]

Furthermore, I mean to reference the fact that bibliographies such as Wright's function in part as arguments for how texts can be combined despite their different associations and relationships. Each act in the process of assembling a collection such as a bibliography requires conscious decisions. In the case of bibliography, these actions can be the choice of institution to visit, collection to browse, information given and recorded in the description, and naturally the parameters for what texts are admitted into the bibliography. The modes of interpretation, the methods, philosophies, and forms in which multiple modes of discrete pieces of information is printed, organized, and understood by both the compilers of that information and the observers of it fit under what I wish to call the rationale of aggregation. The rationale of aggregation is that which defines the operations that inform how to best represent large amounts of information. These operations are subjective, as they involve isolating characteristics of the objects that form a piece of the aggregate though they may not necessarily be presented or understood as such.

This rationale is what drives the process of assembling the assemblage. At the core of this argument will be the examination of enumerative bibliographies, taking as its primary case study Wright's *American Fiction* bibliography but not excluding the contributions of other American literature subject bibliographies. As stated, enumerative bibliographies represent one of the ways in which scholars have attempted to organize and represent aggregated information for use by other scholars. This argument, however, will also extend to the ways in which bibliographies and other collections of texts inform humanities scholarship, especially as scholars have increasingly resorted to digital tools for research. Aggregated collections of texts have become commonplace

7. My definition and conceptualization of *assemblage* is partially inspired by Darnton's communication circuit, which describes the various pathways that influence a text's existence as an object, from author to publisher to reader. Aggregation could be seen as an act which constitutes another "node" on Darnton's circuit. See Robert Darnton, "What Is the History of Books?," *Daedalus* 111, no. 3 (1982). 65–83, ISSN: 0011-5266, accessed April 22, 2017, `http://www.jstor.org/stable/20024803`

in modern literary studies as databases and repositories such as *Early English Books Online*, *Making of America*, and *Eighteenth Century Collections Online*. Digital humanities scholars have led to initiatives to not only use and promote these sorts of collections but also in creating them and interrogating their place as scholarly objects. Scholars originally sheltered in the realms of textual criticism and bibliography have found fresh topics of discussion in the digital possibilities for texts, and so often find themselves confronting the rationale of aggregation when attempting to discuss how databases of texts function as works. Others have taken advantage of the affordances of aggregation and begun to read texts at distance, at a scale far past what is humanly practical. These scholars, however, are not necessarily approaching anything that we have not seen before as scholars. Enumerative bibliographies are early attempts at confronting the rationale of aggregation and in forging a framework for how readers can deal with information as a significant scale; contemporary works in the digital realm that find value in considering the aggregate are informed by the work of the bibliographers that have come before them.

1.1 Data

One of the first considerations for understanding the interpretive possibilities of bibliographic entries and their aggregation is to put them into context with the theories of what it is that makes up bibliographical descriptions: data. My preferred definition of the term is that offered by the Open Archival Information System (OAIS) for its clarity, breadth, and nuance.[8] The OAIS definitions declares data as:

> A reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing. Examples of data include a sequence of bits, a table of numbers, the characters on a page, the recording of sounds made by a person speaking, or a moon rock specimen.[9]

This definition overtly states what are several important concepts for the term as they relate to bibliographical composition. Firstly, the idea of data as ”reinterpretable”, rather than as statically informative, suggests a dynamic value to data. The use of ”interpretitive” again, later on

8. OAIS was adopted as ISO standard 14721 in 2002. Originally a product of the Consultative Committee for Space Data Systems, the OAIS prescribes a system for archival workflows and digital preservation. While I am primarily concerned with the way it has defined the term data and how this may help us approach bibliographical descripton, the OAIS covers a large, interdisciplinary and complex model of archives, preservation, and access. **ccsds˙2002** To reference the ISO standard, see “ISO 14721:2012 - Space data and information transfer systems – Open archival information system (OAIS) – Reference model,” 2012, accessed May 12, 2017, `https://www.iso.org/standard/57284.html`. For the OAIS in a humanities context and its relevance to the archives, see Matthew Kirschenbaum, “The .txtual Condition: Digital Humanities, Born-Digital Archives, and the Future Literary,” *Digital Humanities Quarterly* 007, no. 1 (July 1, 2013), ISSN: 1938-4122

9. **ccsds˙2002**.

in the same sentence emphasizes the process of encountering data as reader-centric, as having its value determined and subsequently defined by the observer. The definition does not suggest data as singularly objective, as possessing a finite amount of truth or factuality, as it assumes reinterpretability as inherent to the objects termed data, and so assumes multiple observers, each with their own means of interpretation that can produce variable outcomes. The word ”processing” to some degree can be understood as nearly synonymous with interpretation, though with a distinction that processing refers to a method of interpretation and re-representation from a technical or computational point of view, rather than that of a human observer. The formalization of data and its ability to be communicated are codependent. In bibliographies, descriptions of titles obey a set sequence of details, determined by the bibliographer, but also often falling in line with a disciplinary consensus: author, title, place, date, with additional details that may be added to help facilitate the particular goals of the bibliography. The data being concordant with other representations of data is meant to allow the reader to both access individual units within the dataset, and see relationships amongst the units. As I will discuss in the New Bibliography section, while the characteristics of data defined here are meant to facilitate the reader's exploration and interpretation of the data, it does not necessarily always succeed in doing so.

The OAIS definition, while it does hint at the unfixed nature of data, bears only a trace of the fact that data itself has undergone interpretation already when it has been sorted, arranged, pared down, etc. In its use of the term ”representation”, the prefix re-, or ”again”, suggests a state prior to that of data. Thus, we may say data is not innocent of human involvement, tampering, or subjectivity, and thus, data itself is not a neutral object, but susceptible to ideology via the the methods and practices that inform the person producing the data. Johanna Drucker has argued for more humanistic approaches to data in her recent work. Drucker has stated in multiple venues that ”data”, derived from the Latin *datum* (”that which is given”), is taken, rather than given, and thus is *not data*, but *capta*.[10] Drucker's aim in the use of the word *capta* to describe what is more commonly know as data as ”always interpreted” as ”no data pre-exists its parameterization.”[11] Parameterization, according to Drucker, is a construction, and the term *capta* opens up the possibility of recognizing and ”acknowledging the constructedness of the categories according to the uses and expectations for which they are put.”[12] The parameterization that Drucker locates as inherent to data/capta is synonymous with the formalization and representation the OAIS definition prescribes.

The way in which data is arranged, according to discrete categories–date, author, title, publica-

10. Johanna Drucker, *Graphesis : visual forms of knowledge production. Johanna Drucker*, MetaLABprojects (Cambridge, Massachusetts : Harvard University Press, 2014., 2014). 128-9, ISBN: 978-0-674-72493-8; Johanna Drucker, “Humanities Approaches to Graphical Display,” 5, no. 1 (2011). 3, accessed February 17, 2017, `http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html`.

11. Drucker, *Graphesis*. 129.

12. Ibid.

tion place, etc.–represent a process of assigning terms or classes to concepts based on an interpretation of a pattern that is observed. For Drucker, these classifications can distort and simplify the complexity of the phenomena being forced into a classification framework, while also erasing the ambiguity amongst the different items arranged and united under the same concept. Drucker's example in this case refers to nations, genders, populations, and time spans; all politically determined and institutionalized concepts that constrain identities to fixed notions, even though they are not "self-evident, stable entities that exist a priori."[13] The same can be said of bibliographical information, whose methods of description are institutionalized according to disciplinary expectations, style and formatting guides of academic journals, and title page printing conventions. Taxonomies that presume authorship as primary and as fitting into a set number of categories (i.e. pseudonym, autonym, anonymous), render the status of author to a position that may not be entirely indicative of the details that accompany the author's relationship to the work. In cases of anonymously published works, the state of a work being known to be by a certain author does not necessarily mean the work itself is not anonymous, as it was published without the name, proclaiming itself as detached from an author. Authorial attributions made to the work are done under the need for the bibliographical information to conform to a standard for reference by the reader of the bibliography. The work itself is filed under a certain protocol and so must be sought according to one's understanding of that protocol.

While it is keen to point out the idea of the "constructedness" of data and its parameters, to take the concept a step further, we must understand how the data signifies–that is, the concept of accuracy as it refers to the idea of trueness or correctness of data–must ultimately also be constructed. If data is meant to accurately depict a phenomenon, and the way the data is arranged is susceptible to subjective interpretation, then the ascription of accuracy must also be a qualitative value informed by interpretation. As will be discussed in a moment, bibliography expresses a desire for accuracy in its descriptions and considerations of material texts. What is termed accuracy however is susceptible to the aims of the scholar compiling the bibliography, who determines the mode of description which then informs how observers of the bibliography can understand information. The bibliographical description must point to the reality of the text, but it is a reality understood by the bibliographer, and assumed by any one who views the bibliography unquestionably. A humanistic approach to data, as Drucker calls for, that is more aware and observant of the variance, subjectivity, and capacities of data to signify more than what may be represented, would also take into account how the accuracy of the data's representation is understood, and by what measures the truth of a text's description corresponds to the text's publishing history, but also how much of the information is appended, by the bibliographer, onto the text with information that exists outside of the textual object.

13. Ibid.

The practice of attempting to classify literature is not innately bibliographical however. Literary works have been cataloged and organized according to their semantic content (that which bibliography attempts to avoid).[14] Genres, in particular, represent a means of understanding collections of texts according to their content as it pertains to following specific themes or stylistic conventions. Wai Chee Dimock, in her introduction to a special issue of *PMLA* on genre describes genres as "fields of knowledge," as she explains:

> Far from being a neat catalog of what exists and what is to come, genres are a vexed attempt to deal with material that might or might not fit into that catalog...The membership–of any genre–is an open rather than closed set, because there is always another instance, another empirical bit of evidence, to be added."[15]

Dimock would seem to agree and anticipate Drucker in her conceptualizing of genre, which drifts into a more scientific rhetoric (empirical, open versus closed set, evidence). Dimock recognizes the problems of attempting to fit genres onto groups of texts. Dimock recognizes the same interpretive nature of data through her thinking about genre: "The spilling over of phenomena from labels stands here as an ever-present likelihood, a challenge to any systemizing claim."[16] Dimock's argument represents a means of understanding aggregation and its effect in the literary realm, but through its relevance to semantic and aesthetic considerations of literature, rather than bibliographic. It is perhaps easier, however, to see how genres and literary traditions, which are more obviously and strongly connected to a literary object and its subjective qualities, represent aggregation efforts that can be restrictive in terms of the claims of what defines a genre, or easily upset when its expectations are not conformed to. Though, what Dimock suggests is that the particular accuracy, or the fidelity, of texts to the genre that is assigned to it, is not the reason for the classification. It is instead, an attempt that can be productive, rather than restrictive if considered in the light that it is only ever going to "vexed", or failing to ever fully encapsulate literature as a whole. The failure on the part of enumerative bibliography to recognize this, and present itself accordingly aware of this fact, lends itself to the modern discussions of databases and assemblage of collections the digital humanities have produced.

---

14. This sentiment is attributed to W. W. Greg: "To the bibliographer the literary contents of a book is irrelevant. This does not mean that special bibliographies should not be compiled, or that the merits of the works included, or somebody's opinion thereon, should not be recorded. It means that this is not the task of the bibliographer." W. W. Greg, "What is Bibliography?," *The Library* TBS-12, no. 1 (January 1, 1913). 46, ISSN: 0024-2160, accessed April 4, 2017, doi:10.1093/libraj/TBS-12.1.39, https://academic.oup.com/library/article-abstract/TBS-12/1/39/953125/What-is-Bibliography

15. Wai Chee Dimock, "Introduction: Genres as Fields of Knowledge," *PMLA* 122, no. 5 (2007). 1378, ISSN: 0030-8129, accessed March 30, 2017, http://www.jstor.org/stable/25501790.

16. Ibid.

1.2 The New Bibliography

To understand how our modern ideals of aggregation manifest, we must first understand how they are in dialogue with, and in some capacity, still at the mercy of the bibliographic principles of the scholars who came before them, primarily the New Bibliographers who discussed and codified the standards for the description of books and their arrangement in enumerated lists. This period of scholarship in the early 20th century systematized the process of book description, collection, and arrangement and prescribed set procedures for how other scholars should be able to view, find, and learn about different texts and works. The process of systematization and prescription was not without its debates, and of particular interest in this section will be the interventions of those who considered in the creation of enumerated lists their interpretive capacity against those who dismissed enumeration as a lower form of bibliography due to its perceived lack of intellectual rigor.

For W. W. Greg, an enumerated bibliography is a list of books described, organized, and compiled according to a "guiding principle."[17] Theodore Besterman offers a modified definition in his history of enumerative and systematic bibliography, preferring, rather than a "guiding principle," a "permanent principle."[18] What the two definitions presuppose with term *principle* is that the bibliography is composed according to a deterministic framework that imposes simultaneously rigidity and clarity unto the items listed, the books described. Rigidity and clarity are codependent in this case, as the conventions of bibliographical description offer a pattern of information representation that creates a comprehensible system for researchers and readers to understand the construction of the bibliography, and therefore use it for its intended purposes.

To more concretely define what is the referent of *principle*, observe the following samples drawn from the listing for Herman Melville's works in the *Epitome* of Jacob Blanck's *Bibliography of American Literature* (*BAL*, compiled by Michael Winship, Philip Eppard, and Rachel Howarthsimple position :[19]

13663. THE WHALE. Lon: Bentley, 1851.
3v. For U.S. edition see next entry.

13664. MOBY-DICK; OR, THE WHALE. NY: Harper & Bros., 1851.
Critical edition published in 1967 (No. 13711).

13666. PIERRE; OR, THE AMBIGUITIES. NY: Harper & Bros., 1852.[20]

17. Greg, "What is Bibliography?" 41.

18. Theodore Besterman, *The beginnings of systematic bibliography*, Burt Franklin bibliography and reference series: #216 (New York : B. Franklin, [1968], 1968). 2.

19. Blanck passed away in December, 1974, before the final three volumes–seven (1983), eight (1990), and nine (1991)–were completed. The final three volumes were completed and edited by others, including Katherine Jarvis, Virginia L. Smyers, and Michael Winship. Winship is credited with the completion in all three volumes.

20. Michael Winship et al., *Epitome of Bibliography of American literature. compiled by Michael Winship, with*

Even from the sample of three listings, it becomes obvious certain patterns that inform the construction of the bibliography, even if they are perhaps unconsciously realized. Not shown is the that the author names are listed alphabetically; Melville's entries appear between authors Cornelius Mathews (1817-1889) and Joaquin Miller (1837-1913). The author's name assumes the primary position for the bibliography's organization, and thus is the first layer of an information hierarchy or taxonomy for conceiving of the texts listed. Beyond the author's name, we find the titles of the texts included, with the accompanying information of place of publication, publisher, and year. It is at the year that we see the next organizational pattern–chronological–emerge, as the end of the description provides us with a means of knowing why the texts are listed in the order they are. The entry for *Pierre* (1852) at the end of this extract is what provides us evidence that the first two entries referring to two separate editions of *Moby-Dick* are in chronological order despite both listing 1851 as their publication year. In the case of Melville, this informational structure is useful in elucidating a basic publication details about *Moby-Dick*: it's original title was not *Moby-Dick*, but in fact *The Whale*, the title under which it appeared first in London, England before its print run in America. The additional details provide further information, letting us know that while the British edition was first, it was the American that formed the basis of the first critical edition, described later in Melville's entry.

We can clearly label, then, the supposed principles–guiding, permanent, or otherwise–Winship, Eppard, and Howarth assumed in their adaptation and interpretation of Blanck's bibliography.[21] The *Epitome*, like the *BAL*, privileges the author as the prime piece of information that forms a bibliographic entry. In doing so, it assumes, like any bibliography whose construction is primarily modeled by an alphabetical author listing, that the primary use of the bibliography will be in researching individual authors; that the first means of reference for the scholar is to locate the author whom has been judged responsible for a given text and proceed from there in other directions (i.e. referencing other authors or other texts within the author listing). The secondary characteristic, the

Philip B. Eppard and Rachel J. Howarth (Golden, Colo. : North American Press, 1995., 1995). 200-1, ISBN: 978-1-55591-950-4.

21. Winship, Eppard, and Howarth explain in the introduction to the *Epitome* the principles they assumed for their work in adapting and interpreting Blanck's bibliography:

> "Our goal has been to provide a useful complement to the full *BAL* rather than a replacement for it. We have followed the scope and style as set forth in Blanck's "Preface" in the first volume of *BAL* and have limited ourselves to information contained in the published volumes. In particular we have not included editions of an author's works that have appeared since the publication of that author's list in *BAL*, nor have we incorporated the few corrections or additions that have been discovered since publication of the original volumes."Winship et al., *Epitome of Bibliography of American literature. compiled by Michael Winship, with Philip B. Eppard and Rachel J. Howarth*. vi

The primary criticism one could leverage against this practice as a principle is that it does not turn the *Epitome* into a reference that could point the reader to corrections over Blanck's errors, and would rather have the reader led to Blanck's entries by their own virtue, rather than seek to modify access to individual texts that Blanck does not describe, whether accurately or at all. In short, the *Epitome* in this case represents bibliography for bibliography's sake.

chronological organization, is subservient to the primacy of the author. The chronological listing is framed by the author and so places the author's texts within a timeline, but only with each other text under the author's individual portion of the list. Due to the isolating effects of consigning titles under an author header, more difficulty, on the part of both the bibliographers and the readers, becomes apparent when attempting to obtain information that is not determined by authorial association.

The distinction between Greg's use of *guiding* and Besterman's *permanent*, however, introduces complications into this supposition of how one constructs, and ultimately, obeys the principles the bibliographer lays out. For Besterman, *permanent* indicates a sort of finity, constraint, and superiority on the part of the bibliographer compiling the bibliography that a term such as *guiding* does not. What Besterman refers to with his idea of permanence is that which defines the predetermined knowledge the compositor of the bibliography comes into the project possessing. This model of bibliography, that which is derived from what is "known" beforehand, demonstrates a disposition found among twentieth-century bibliography that imagines the discipline as a science, carrying the connotation of laws and facts, which are presumed to be stable, infallible, and observable (without considering who the observer is). When discussing St. Jerome's *De Viris Illustribus* (fourth century CE), which Besterman deems an early example of systematic bibliography, albeit accidental, he claims Jerome "looked upon his compilation as a piece of theological propaganda. He did not put out his bibliography to to guide or to instruct, but to convert."[22] Besterman reveals not just St. Jerome's predispositions here in terms of what a list of texts *should* do, but his own as well in clarifying Jerome's motives. Besterman believes bibliographies should "guide" or "instruct", which in itself reveals the placement of bibliographers as curators of facts, as those who lead others to knowledge by which they can be informed according to the bibliographer's standards.

Besterman's definition demonstrates a symptom of the New Bibliography in terms its approach to the description, enumeration, and analysis of texts.[23] Other scholars have explored the premise that New Bibliography conceived of itself as scientific in nature and constructed the field in ac-

22. Theodore Besterman, *The beginnings of systematic bibliography*. 9.

23. The *Oxford Companion to the Book*'s entry for "New Bibliography" goes as follows:

> The New Bibliography involved 'the application of physical evidence to textual problems' (Tanselle) and one of its key achievements has been its systematic and rigorous methodology for describing such physical evidence...It's most enduring (and contentious) legacy has been in the field of editorial theory, where its intentionalist and eclectic editorial principles (refined first by Greg, later by Bowers, and more recently by Tanselle) long dominated the production of critical literary editions.

Michael F. Suarez and H. R. Woudhuysen, *The Oxford companion to the book. edited by Michael F. Suarez and H.R. Woudhuysen* (Oxford ; New York : Oxford University Press, 2010., 2010). 963, ISBN: 978-0-19-860653-6. G. Thomas Tanselle, "Bibliographical History as a Field of Study," *Studies in Bibliography* 41 (1988). 40, ISSN: 0081-7600, accessed April 23, 2017, `http://www.jstor.org/stable/40371876`

cordance with mechanical and technical methodologies.[24] New Bibliographers are not hesitant to respond this claim affirmatively, though. As Greg notes:

> ”Facts are observed and catalogued by the systematizers, and then suddenly, as if by chance, an idea is born that introduces order and logic into what was the mere chaos, and we are in possession of a guiding principle, of an instrument of thought and investigation, that may transform the whole of our relation to knowledge or alter the face of the physical globe.”[25]

Bibliographers certainly belong to the same population as those Greg calls ”systematizers.” It is the responsibility of the bibliographer to order facts according to those ”guiding principles” which in turn produce knowledge. For Greg, knowledge seems to only come out of ”order and logic” once it has been applied to ”facts”, or what may be termed in other circumstances as ”raw data.” Greg, however, seems to consider enumerative bibliographies to be ”raw data” that have not yet had ”order and logic” applied to them, and so are not necessarily productions of knowledge themselves, but only aids to its manufacture. Philip Gaskell is of a similar opinion; in his *New Introduction to Bibliography*, appearing almost sixty years after Greg's above statement, Gaskell understates enumerative bibliographies as useful, but not the ”purpose” of bibliography. Instead, their job is to aid in the study of literature as reference tools and to aid bibliography in deterministically proliferating ”accurate” texts.[26] A. S. G. Edwards makes a similar claim in discussing enumerative bibliographies: ”One accepts, I assume, that the aim of any enumerative bibliography is to achieve as close an approximation to definitiveness as is practicable.”[27] Terms such as ”accuracy”

24. G. Thomas Tanselle has published a lengthy article on the subject of bibliography and science, where he revisits the debate. He argues that science was only ever an analogy to refer to the empirical or systematic nature of bibliography and editing and to contrast it with the bibliophilic tradition that preceded the New Bibliography. Tanselle does, however, also lay some of the blame for the debate on bibliographers themselves for continued use of scientific terminology. G. Thomas Tanselle, “Bibliography and Science,” *Studies in Bibliography* 27 (1974). 57, ISSN: 0081-7600, accessed April 27, 2017, `http://www.jstor.org/stable/40371588`. More recently, some scholars have also approached the debate by discussing it as a symptom of the cultural moment the New Bibliography emerged. Amanda Gailey discusses the Greg-Bowers method and the scientific rhetorical grounding of it as a response to the Cold War and the need to compete with scientific fields for federal funding and prestige. Amanda Gailey, *Proofs of Genius: Collected Editions from the American Revolution to the Digital Age* (2015), `http://hdl.handle.net/2027/spo.13607061.0001.001`. Amy E. Earhart has tied the Greg-Bowers method and its focus on the ”purity” and ”corruption” of the text to both Bowers' interest in dog breeding and the resistance of textual scholarship to diversity issues. Amy E. Earhart, *Traces of the old, uses of the new : the emergence of digital literary studies. Amy E. Earhart*, Editorial theory and literary criticism (Ann Arbor : University of Michigan Press, [2015]., 2015). 36-7, ISBN: 978-0-472-07278-1. These discussions are all almost entirely dedicated to discussing the debate in regards to textual criticism and editing, however, and not necessarily covering biliography as whole and its other subsidiary methods of description or enumeration.

25. Greg, “What is Bibliography?” 41.

26. Philip Gaskell, *A new introduction to bibliography. Philip Gaskell* (New York ; Oxford : Oxford University Press, 1972., 1972), ISBN: 978-0-19-818150-7.

27. A. S. G. Edwards, “Some Problems in Modern Enumerative Bibliography,” *Text* 1 (1981). 331, ISSN: 0736-3974, accessed January 14, 2017, `http://www.jstor.org.ezproxy.library.tamu.edu/stable/30234257`.

or ”definitive” depict a particular stance towards information that belies the ultimately subjective nature of texts and their production. Unironic or unqualified use of the word ”accuracy”, or even its corresponding term–precision–forgets that in the case of texts, what is ”accurate” is ultimately subject to and defined by the whim of the bibliographer themselves.

Fredson Bowers is conscious of this detail when he attempts to conceive of the place of enumerative bibliographies–or, to him, catalogues, handlists, or checklists, terms which would likely not stand up to the scrutiny of formal librarians. He, like Gaskell, understates enumerative bibliographies by making them subservient to descriptive bibliographies:

> Their primary purpose is to make available a listing of books in a certain collection or library, or else in a certain field, such as a specific period, a particular type of literature, a definite subject, or an individual author. Noting the existence of these books is the end-all and be-all of a catalogue, and under ordinary circumstances only the minimum of identifying details is provided, as author, title (abbreviated when necessary), the date and possibly the place of publication, and occasionally the format. Some catalogues may include the name of the printer or publisher, or both. The writer may compile his list partly from other catalogues and partly by personal examination of the books, supplemented by notes furnished by contributing libraries or scholars; but except in extraordinary cases he is not concerned with the textual history, circumstances of printing, or variation within issue (sometimes even within edition) of the books listed.[28]

What Bowers does note, however, is that ultimately, these mere lists are subjective. While disparaging enumerative bibliographies whose ”end-all and be-all” is simply stating their existence, he notes that details included in the lists are not always absolute. The compiler of the list is in control of what information is provided to the reader, and, in effect, can determine how the list is used as a research and reference tool. Bowers perhaps is seeing the nuance of Greg’s ”guiding principle” that Besterman’s ”permanent principle” discarded. To Greg’s credit, he saw bibliography as empirical, based on evidence, but not rationalistic. In the ironically titled ”Rationale of the Copy-Text”, Greg argues that the choice of a parent text which guides the creation of a new edition by an editor be informed by the most ”authoritative” copy of the text that can be located (by means, most likely, of an enumerative bibliography that goes uncredited). But Greg carefully notes that ”authority” is always relative, never absolute.[29] For Greg, expertise–formal training and education in a subject–allows for the ability to determine ”authority.” But Greg does not necessarily turn his

28. Fredson Bowers, *Principles of bibliographical description*, St. Paul’s bibliographies: 15 (Winchester, U.K. : St. Paul’s Bibliographies ; New Castle, Del. : Oak Knoll Press, 2005., 2005). 3, ISBN: 978-1-884718-00-7.

29. W. W. Greg, “The Rationale of Copy-Text,” *Studies in Bibliography* 3 (1950). 19–36, ISSN: 0081-7600, accessed April 24, 2017, `http://www.jstor.org.ezproxy.library.tamu.edu/stable/40381874`.

qualification of subjectivity onto himself. David F. Foxon, however, is apt to do so, as he targets both Bowers and Greg: "My researches suggest that some at least of our accepted conventions result from the idiosyncrasies of individual scholars; these were uncritically adopted by others and have finally come to be regarded as scientific." [30]

Foxon's criticism, again like most New Bibliograpy discussions, is focused on description, as his primary example in this case refers to Greg and Bowers' apparent disagreement about labeling recto and verso pages of a printed leaf and its adaptation into collation formulas despite the convention being, to Foxon, obviously illogical.[31] Foxon's critique is not irrelevant to enumeration, however. Certain standards as to the arrangement of bibliographies are in place that obfuscate the interpretive nature of aggregation and enumeration. Alfred W. Pollard lists three primary methods of organization that are prevalent in bibliographies:[32]

(i) Alphabetical

(ii) Chronological

(iii) Logical

Of relevance to this discussion are the first two points, which correspond to the major American literature bibliographies, including Wright.[33] While stated in 1907, these methods of classification, present long before the New Bibliography, and enduring long after its heyday and the emergence of critical theory, have affected the way in which information is presented and, thus, interpreted. The proximity of pieces of information suggests a relationship. Arundell Esdaile acknowledges this when he lambasts the concept of alphabetical organization:

30. David F. Foxon, *Thoughts on the history and future of bibliographical description* (Berkeley, CA: School of Library Service, University of California, 1970). 7 In a memorial essay on Foxon, James McLaverty claims that Foxon's criticism against Greg, Bowers, and the enshrined practices of bibliography were derived from two sources: first, the development of technology in the 20th century, specifically the ability for scholars to xerox pages and compare them for both analysis and description. And second, that traditional methods were flawed in their ability to differentiate among editions.James McLaverty, "David Foxon, Humanist Bibliographer," *Studies in Bibliography* 54 (2001). 103, ISSN: 0081-7600, accessed April 25, 2017, `http://www.jstor.org/stable/40372245`

31. To explain further, the convention Foxon is discussing holds that one should only ever explicitly reference verso pages, with either a *b* or *v* while recto pages go unremarked. Foxon addresses the fact the Greg and Bowers agree on the method but not its meaning. The lack of a marker that signifies recto or verso pages on a leaf, to Greg, says that the entire leaf should be considered, rather than a page, but Bowers claims that leaving the signature unmarked would suggest recto unless context implied differently. This moment of disagreement points to ambiguity that Foxon claims is antithetical to a scientific system. David F. Foxon, *Thoughts on the history and future of bibliographical description*. 8-9

32. Alfred W. Pollard, "The Arrangment of Bibliographies," in *Alfred William Pollard : a selection of his essays. compiled by Fred W. Roper*, The Great bibliographers series: no. 2 (Metuchen, N.J. : Scarecrow Press, 1976., 1976). 133, ISBN: 978-0-8108-0958-1.

33. Not discussed here is what Pollard means by the term "logical", where he considers the "natural sequence" of a subject. An example here is a bibliography of mathematics, which Pollard states would have to be subdivided into smaller sub-topics (arithmetic, algebra, etc.) to be comprehensible and useful. ibid. 137

> Some bibliographers have simply sorted the titles into the alphabetical order of authors; but that is mere intellectual laziness or want of imagination (perhaps the same thing); for while the alphabet enables the searcher to get access in a library to a particular book of whose existence he is aware, or, it may be, to refresh his memory as to a title or date, or other detail in the title, it serves no other purpose. The alphabet does nothing to collocate material bearing on the same or a closely allied side of the subject; it serves you up impartially the prunes and prisms together.[34]

Esdaile's mantra for the use of referential materials and bibliographies specifically is that they should be "illuminating" in their organizational method for the researcher.[35] The bibliographer's task and thinking should be on its organizational structure, as failure or indeterminacy in this area would mean the reader of the work, the student and would-be researcher, would "lose his way."[36] Esdaile's tone is patriarchal and authoritative, not dissimilar to that of Besterman's, in that both place the bibliographer in a superior position over the reader. Esdaile does, however, hold the bibliographer accountable for their organization should it present poorly aggregated material that does not allow space for interpretation via its combination of bibliographic descriptions. The fundamental aspect of enumeration is to ensure an open-ended but well-mediated pathway between the subject of the bibliography and its reader. *Collocation* is a favored term in Esdaile's manual as it hints at the interpretive nature of proximity. Esdaile's use of the term references patterns that should, ideally, be easily perceivable or possibly emerge when one entry is compared amongst many others, seemingly in the same physical space of a page, section, or chapter, as physical separation caused by a critically detached organizational scheme, such as alphabetical listing, hinders the reader's ability to recognize latent patterns.

Esdaile resonates with Pollard, who also considers alphabetic listing as unfavorable. On the one hand, alphabetical order is, according to Pollard, the most fundamental organizational system that one can assume a literate person will recognize; it benefits from being inherent to the ability to read, and thus is the simplest. This, though is precisely the reason why Esdaile declares it lazy and serving "impartially the prune and prisms together." Alphabetic organization is impartial because it is detached from the subject of bibliography, and bearing no relation to the connections, in the case of literary bibliographies, among authors–the primary subject of alphabetization. A relationship inferred between Melville and his colleagues in the *Epitome*, Cornelius Mathews and Joaquin Miller, purely on the basis of their proximity in an alphabetical listing would of course be erroneous. Any information that could emerge is at the mercy of chance. More consistently, research

34. Arundell James Kennedy Esdaile, *A student's manual of bibliography. [3d rev. ed.]. by Arundell Esdaile ; revised by Roy Stokes*, The Library Association series of library manual, 1 (New York : Barnes & Noble, [1954], 1954). 364.
35. Ibid. 35.
36. Ibid. 20.

would require further digging into the entries for other details–year, publisher, collation, etc.–to form a more logically sound thesis. Thus, the the complexity of the bibliography's organization is increased without substantial aid to the information it provides[37] The simplicity, though, is why Esdaile also claims it is so widely used; it's ease of access for the reader to grasp, despite the fact that it offers no new information and produces no knowledge. The "prunes and prisms" Esdaile mentions describe his ultimate opinion on the practice: it looks good, or has an aesthetic quality, but nothing more.

It is worth mentioning as well that when Pollard lays out his three common categories, he lists alphabetical as a "according to the names of authors."[38] Traditional print bibliographies often organize themselves in such as to allow the author to occupy the prime position, both in their overall organization scheme and at the level of the individual description. The *BAL* is entirely guided by the author as a central figure for the way it presents its information and how it divides its volumes, while others like Wright's *American Fiction* is divided into volumes by chronology (1774-1850, 1851-1875, and 1876-1900), but within each volume the listed titles are arranged alphabetically by author. While Pollard and Esdaile do criticize alphabetical listings, they both let the concept of the author as first and foremost in a description go with little comment. Pollard does claim the "reader who already knows the book which he wants will be able to find it at once under the name of its authors," but this only hints at the root of the issue of the arrangement: a reliance upon the assumptions that guide the bibliography.[39] The primary of the assumptions being that information is sought based on a larger, overarching questions that presumes the author or name listed as relevant to the conversation, when it is not necessarily always the case. Nineteenth-century American publishing, of course, did not lack its share of anonymous and pseudonymous titles. An alphabetical list according to author is immediately troubled by this fact. Wright's first volume of *American Fiction* (1774-1850) begins with an anonymous entry:

> AN ACCOUNT OF THE MARVELOUS DOINGS of Prince Alcohol, as Seen by One of His Enemies, in Dreams. [N.p.] 1847. 72 p. 12mo
>
> NYP[40]

When an author is absent, Wright naturally defaults to a title, but this not only disrupts the inherent order of the bibliography Wright wished to enforce, but also serves to bury anonymous titles

37. An exception is in the case of family relationships where patronymics remain static. For example, Amos Bronson Alcott and Louisa May Alcott are together in the *Epitome*. A relationship could be inferred by a hypothetical unfamiliar scholar, though it would require confirmation. This is seemingly the only instance where alphabetization does help the reader understand a relationship between entries, but the limited application does not justify the lack of coherence amongst the other 279 authors in the *Epitome*.

38. Pollard, "The Arrangment of Bibliographies." 133.

39. Ibid. 134.

40. Lyle Henry Wright, *American fiction, 1774-1850 ; a contribution toward a bibliography. by Lyle H. Wright*, Huntington Library publications (San Marino, Calif., [publisher not identified] ;, 1939., 1939). 1.

in odd places that makes tracing them more difficult, especially when the prospect of shortened titles, as Wright mentions he did casually in the preface, are considered.[41] Further complicating the scheme, the listing is interjected with notes that reference known pseudonyms of authors. When looking for Mark Twain in the second volume (1851-1870), readers will find in between title entries a note to find the works sought after under Samuel Langhorne Clemens.[42] Similarly, searching for Fanny Fern will direct the reader to instead look under Sarah Payson Willis Parton.[43] *American Fiction* is not wholly consistent, however, as it does include Harriet Jacobs' *Incidents in the Life of Slave Girl*, but under the pseudonymous Jacobs with only a passing reference to the autonymous Linda Brent. Wright provides the following odd description:

> [JACOBS, MRS. HARRIET (BRENT).] Incidents in the Life of a Slave Girl. Written by Herself... Edited by L. Maria Child. Boston: published for the author, 1861. 306 p.
>
> AAS, BP, H, HEH, LC[44]

The decisions for how authors are designated in the bibliographical descriptions represent a moment where the principles of aggregation, here inclined towards authorship and alphabetization, produce friction against the concept of bibliographies seeking to produce accurate and definitive lists. On the hand, the preference of Wright, and other bibliographers, on autonymous author entries for their listing is on the one hand deferential to the work of the person who created the work. The practice, however, also represents a process of interpretation of data that runs counter to more dominant discourse of these authors by literary scholars who are supposedly meant to be served by the bibliography. Mark Twain, rather than Samuel Clemens, of course, is the dominant name attached to the works of *Huckleberry Finn*, *Connecticut Yankee*, and so on. The same can also be said of the less canonical authors such as Fanny Fern, the name attached to even the most recent editions of her works[45] Bibliographical descriptions of the texts that abandon the common discourse of author reference offer a competing claim as to the creator of the work. This is an act

41. Ibid. ix.

42. Lyle Henry Wright, *American fiction, 1851-1875 : a contribution toward a bibliography. Lyle Henry Wright*, Henry E. Huntington Library and Art Gallery Publications (San Marino, Calif. : Huntington Library, 1957., 1957). 341.

43. ibid. 120. Wright does, however, denote names absent from the title page with square brackets ([]), which does include autonymous names that supplant pseudonyms; his physically places distance between the title of the work and the listed author name in this case. Blanck in the *BAL* also defaults to autonyms for the authors he lists, though there is little to help the reader realize this. Mark Twain is found under Samuel Clemens in volume 2. This introduces some difficulty for the reader at a practical level who may not go into the bibliography with the knowledge of Blanck's organization, and so goes searching volume 8 for Twain, rather than the correct volume.

44. ibid. 179 More about the oddities of the inclusion and description of Harriet Jacobs in a bibliography of fiction will be explored in a future chapter.

45. See Susan Belasco's edition of *Ruth Hall*. Fanny Fern, *Ruth Hall: A Domestic Tale of the Present TIme*, Reprint edition, ed. Susan Belasco (New York: Penguin Classics, February 1, 1997), ISBN: 978-0-14-043640-2

interpretation. Wright inherently recognizes this by providing the signposts in his bibliography that point researchers from the pseudonyms, which are assumed to be sought first, and their place in the listing to Wright's preferred method of classification–the autonym.

This issue, I would argue, is bound to the New Bibliographical desire for accuracy and definitiveness, but shows that these concepts are subjective when it comes to attempting to codify and arrange a subject such as early American writing. Wright's bibliography demonstrates an ideal that understands the author as synonymous with the person writing and publishing the work, rather than observing and adhering to what the text claims about its author. This makes sense in cases such as Washington Irving, whom is used as the authorial reference in regards to his works written pseudonymously–*The Sketchbook*, for example, composed by Geoffrey Crayon, or Diedrich Knickerbocker's *History of New York*–wherein the author figures are themselves constructed as fictional characters.[46]. However, in the cases of Twain and Fern, these issues are less clear. Neither Fern nor Twain are offered as fictional characters by the texts that bear their names. The writings of Fern, including her earliest published monograph, *Fern Leaves from Fanny's Portfolio* (1853), does not suppose Fern to be a fictional entity, but instead a different means of referencing the author that provides a circumstantial detachment from the author's legal name. The distinction, however, is that Fern and Twain adopt the literary status their counterparts, Clemens and Parton, leave behind in the course of both their publishing and circulation, wherein the names printed alongside the text become inherent to how the text is received and read. Scholars working with the texts continue to hold Twain, rather than Clemens, as the author figure. Thus, when *American Fiction* demotes Twain and Fern to pseudonym while promoting Clemens and Parton to author figure in its organization, it does so in contest of the dominant the discourse. Instead, it has located its concept of accuracy in a discourse outside of the realm of the subject it is responsible for enumerating.

When a bibliography is running in opposition to the reader, it ends up making the process of turning data into information more difficult. Pollard prefers, over the alphabetical listing that, in conjunction with a fixation on the author figure, complicate knowledge-seeking and "illumination" (to use Esdaile's word), the chronological format such as that used by Wright as the dividing line between the three volumes of *American Fiction*. For Pollard, chronology is easily comprehensible, similar to the alphabet in terms of the capabilities of the assumed reader, but is in general more generative and expressive as a organizing principle. Wright covers the years 1774 to 1900 in *American Fiction*, with each volume containing a division of that timeframe: 1774-1850, 1851-1875, and 1876-1900. This timeline is seemingly arbitrary to the reader, as Wright does not offer comment as to the *why* of his demarcation. The most Wright provides is some meditation as to the what he perceives as the general overarching themes of the books included in the bibliography.

46. In the works of Irving mentioned, Knickerbocker and Crayon are narrators, but are also telling stories about themselves, implicating their own fictionality in the course of the narratives.

From the preface of volume 2 (1850-1876):

> The momentous events that occurred during this quarter century are reflected in the fiction of the period. The slavery question, pro and con, was the theme of scores of novels, and as many more covered the Civil War, a national catastrophe that induced authors to attempt to be more realistic in their writing. The westward flow of the population was not overlooked...During the 1850's the sentimental novel reached its peak in popularity, aided and abetted by the large increase in women writers. And the woman's rights movement gained impetus through the numerous novels and short stories which presented it in a sympathetic vein. Religion, including controversies between denominations, was also a favorite subject with authors.[47]

One need not get too far in the previous volume to see the porous nature of the themes Wright identifies for the second. Recall the first anonymous entry, *An Account of the Marvelous Adventures of Prince Alcohol* by "One of His Enemies," a temperance novel addressing the religious controversies Wright describes.[48] Similarly, an 1849 novel, *Amelia Sherwood; or, Bloody Scenes at the California Gold Mines* references westward expansion.[49] J. Elizabeth Jones' *The Young Abolitionists; or, Conversations on Slavery* (1848), published by the Boston Anti-Slavery Office is obviously in dialogue with the "slavery question, pro and con."[50] Since the only apparent justification for Wright's chronological division is his thematic considerations, it is worth addressing how that division in the case of American history here, and of chronological classification systems generally, are to some degree always arbitrary and reliant upon judgment. Wright's thematic reasoning offered in the prefaces of his three volumes provide a general sense of contents, but do not clearly elucidate with specificity the nature of the time period bound by each volume.

Rationalizations attempting to explain Wright's demarcations meet inevitable walls in their logic. The second volume beginning at 1851 breaks from F. O. Matthiessen's pinpointing of 1850 as the first year of his "American Rennaissance" (1850-1855).[51] The 1851 mark Wright draws causes the individual author bibliographies to become split, with works by Hawthorne and Melville, included in all three volumes.[52] Book history complicates Wright's chronology significantly by moving events and innovations with drastic effects on American publishing earlier than 1851. Copyright (1790 with revisions in 1802 and 1831), the postal service (1792), continental

47. Wright, *American fiction, 1851-1875*. vii.
48. Wright, *American fiction, 1774-1850 ; a contribution toward a bibliography. by Lyle H. Wright*. 1.
49. Ibid. 5.
50. Ibid. 156.
51. Matthiessen first published *American Rennaissance* in 1941. Wright could have known Matthiessen's work by the time the second volume was being compiled. F. O. Matthiessen, *American Renaissance: Art and Expression in the Age of Emerson and Whitman* (New York ; London: Oxford University Press, 1980)
52. Posthumous works by Hawthorne, who died in 1864, are located in the third volume. Both *Doctor Grimshawe's Secret* (1883) and *The Dolliver Romance, and Other Pieces* (1876) are described there.

railroad (1869), telegraph (1844 with its transcontinental implementation occurring in 1862), and dominance of the machine press over the hand press by the 1840s shift bibliographical qualities of texts outside of Wright's own parameters.[53] The bibliography's chronological ordering principally based on theme and disconnected from a more specific historical perspective means that Wright has taken the most interpretive step in organizing the publication dates of the titles he lists. This puts a constraint upon the researcher, whose ability to evaluate events, phenomena, writers, or subjects that span years outside the scope of single bibliography find themselves facing, at a practical level, more difficulty in searching and comparing the information Wright offers than those who find their questions approachable with a single volume.[54]

Lastly, in reference to the composition of a bibliography that is outside of Pollard and Esdaile's consideration of the arrangement, but inherent to it are the selection principles of a bibliography and whether or not the reader is made aware of the biases of the compositor (or the institutions from which they get their information). Again, Wright is a useful example as one whose goal of compiling a list of early American fiction is not as fundamentally clear as it may first appear. In the preface for each volume, Wright presents a similar worded statement in an attempt to clarify his aggregation process: "The design of this bibliography is to list the American editions of novels, novelettes, tales, romances, short stories, and allegories, in prose, written by Americans."[55] This statement is not as all-encompassing as a reader may suppose, however; Wright excludes, despite their relevance to the above system: "annuals and gift books, publications of the American Tract Society and the Sunday School Union, juveniles, fictitious Indian captivities, jestbooks, folklore, collections of anecdotes, periodicals, and extra numbers of periodicals," as well as essays.[56] Wright is honest in what this may mean for his entries, as he notes that his parameters cause some questionable exclusions even for canonical authors; Poe's "The Balloon Hoax" (1844) and Whit-

53. Robert A. Gross and Mary Kelley, eds., *A History of the Book in America: Volume 2, An Extensive Republic, Print Culture, and Society in the New Nation, 1790-1840*, vol. 2 (Chapel Hill, NC: University of North Carolina Press, 2010) See essays by Meredith McGill on copyright and the postal service; Richard R. John for a discussion of early postal service operations, and their relationship to the stagecoach industry; and Robert A. Gross' introduction for discussions of the telegraph and railroads. See Scott E. Casper's introduction where describes the gigantic shift in the amount of railroad coverage began in the 1840s. Scott E. Casper et al., eds., *A History of the Book in America: Volume 3, The Industrial Book, 1840-1880*, vol. 3 (Chapel Hill, NC: University of North Carolina Press, 2007). See Jen A. Huntley-Smith, "Print Cultures in the American West," in *Perspectives on American Book History: Artifact and Commentary*, Perspectives on Book History (Amherst and Boston: University of Massachusetts Press, 2002). 255–284 for a discussion of book history in the American west as result of the telegraph and railroad.

54. Amusingly enough, Wright's division of the three volumes does show an interesting pattern. Volume 1 contains just under 2200 entries (Wright does not provide exact numbering, but an estimate); volume 2 contains 2832 numbered titles; and volume 3 enumerates 6175 titles. Physically, each subsequent volume is larger than the other, visually suggesting that as time moved forward, more and more American titles were published. This a base assumption that could be more complicated; there may be better preservation efforts for more recent titles, or they have had easier times surviving. The institutions Wright traveled to may have privileged later titles over earlier ones, generally. Or, Wright's own standards and selection principles grew more relaxed.

55. Wright, *American fiction, 1774-1850 ; a contribution toward a bibliography. by Lyle H. Wright*. ix.

56. Ibid.

man's *Franklin Evans* (his 1842 temperance novel) are consciously excluded because of Wright's decision to omit extra numbers of periodicals.[57] While some of his omissions seem to suggest themselves as obviously outside the bounds of fiction (essays and jestbooks). Others raise more questions or are in need of an argument that is not provided.

The categories of "juvenile" or "fictitious Indian captivities" raise a question as to why these would not be desirable in Wright's bibliography, and its effects can be clearly noted when browsing the listings. The exclusion of juveniles, for example, results in Louisa May Alcott's listing the second volume to be curiously empty. Wright lists for Alcott five titles: *Hospital Sketches* (1863), *Hospital Sketches and Camp and Fireside Stories* (1869), *Moods* (1865), *On Picket Duty, and Other Tales* (1864), and *Work: A Story of Experience* (1873). Absent are the arguably more popular and relevant works for reference to Alcott: *Little Women* (1868), and its sequels *Little Men* (1871) and *Jo's Boys* (1886, this text would be located in the third volume were it included). Wright's judgment, then, considers Alcott's major works as outside the bounds of more "adult" literature that he wishes to enumerate.[58] This is, again, an interpretive judgment that offers an argument both to claims as to what is valuable within the realm of American publishing and thus what should be accessed by researchers. A genre such as the "juvenile" is less stable than the deterministic methods of arranging and presenting bibliographies acknowledge, in so far the methods operate on a completely binary system of either including or excluding an item in their arrangement.

Wright's other major oversight here is the omission of periodicals. The expansion of communication technologies and the industrialized machine press gave rise to serial novels as a prevalent form of fiction and literary writing that original embodied such works as Edgar Allen Poe's *Arthur Gordon Pym* (1838), Melville's *Israel Potter* (1855), and even Harriet Beecher Stowe's *Uncle Tom's Cabin* (1852). These works would eventually come to be found in book form that allowed them to be noted and described by Wright. Other serial publications were left out of this, however, including prominent works by African-American authors, such as Martin Delany's *Blake, or, the Huts of America* (serialized in the *Anglo-African Magazine* in 1859), which is left undescribed by Wright. By ignoring serials, Wright also coincidentally ignores a prominent venue for black writers in the nineteenth century. As Eric Gardner has stated, black serials were "*the* central publication outlet for many black writers–and especially for texts that were *not* slave narratives."[59] Additional casualties of Wright's method include Julia C. Collins' *The Curse of Caste; or, The Slave Bride* (1865) as well as Frances Harper's *Minnie's Sacrifice* (1869).

57. As reported by Wright, "The Balloon Hoax" was published in *The Extra Sun*, April 13, 1844. *Franklin Evans* appeared in *The New World*, II, No. 10, Extra Ser., No. 34.

58. "Adult" becomes a term Wright adds to his preface in future volumes and revised editions.Wright, *American fiction, 1851-1875*. vii

59. 10 Eric Gardner, *Unexpected Places* (University Press of Mississippi, 2009). emphasis Gardner's, ISBN: 978-1-60473-284-9, accessed May 16, 2017, `http://muse.jhu.edu/book/9988`.

On the other hand, sometimes Wright becomes too inclusive with African-American literature, and offers in a bibliography of fiction, autobiographical works such as the previously mentioned Jacobs' *Incidents in the Life of the Slave Girl*, and Solomon Northup's *Twelve Years a Slave* (1853). In addition to the exclusion of serials which barred black authors from receiving more descriptions, the inclusions also present problems when they become described and framed under an erroneous assumption that then paints how the reader will read the work. With the overarching claim that the texts represent fiction, slave narratives such as Jacobs' and Northup's have their receptions altered in a way the discredits their experiences and stories. At the same time, their inclusion possesses the possibility of creating an errant scholar whose reading of those texts pulls them out of their historical context, or supplants autobiographical content and lived experience with aesthetic and literary motives. Cases such as Jacobs and Northup are easy to determine were this happening, given their recent prominence in literary studies. It is for the titles that have not yet received the same consideration as these that are in vulnerable positions.

Several other oddities exist within the listings of Wright beyond the omissions and the gaps they produce. An American edition of Charles Dickens' *The Mystery of Edwin Drood*, is described; this edition of a work Dickens did not complete before he died in 1870 was completed by an American, Thomas Power James, and so to Wright seemed to belong amongst the collection of nineteenth- century American fiction more so than texts Wright excludes.[60] The issues of Wright and the guiding principles of his arrangement and presentation, and thus, their constraints as I have discussed here–their alphabetization, chronology, and selection principles–become a more visible issue when we consider the *Wright American Fiction Project*, which has only provided information derived from the second volume of Wright. The *Wright American Fiction Project* takes Wright's presumptions, assumptions, and overall methods at his word, and does little in regards to questioning his selections, but does make editorial decisions in regards to the texts Wright lists in order to build their repository and its functionality. When presented as a digital resource, a dataset for distant reading, and a guide to early American fiction, the limitations and constraints of Wright become more prominent and their effect is felt far more strongly once the issues begin to resist or conflict with aggregate unity of the bibliography.

1.3 Databases, Datasets, and Aggregation

With the emergence of digital humanities and its byproducts, scale and aggregation become a more prevalent force for scholars outside of the rigid listings of decades-old bibliographies. The functionality of these digital resources not simply conceptualized as simple listing or guidance to subject areas, but also in their ability to be viewed, used, and thought about at a scale beyond

60. Wright, *American fiction, 1851-1875*. 422.

the individual book. Digital repositories of texts do not face the same considerations enumerative bibliographies have in terms of their limitations (for both good and bad). Databases and datasets produced are not simply categorized as mere finding aids or reference guides, but as cohesive objects that unite discrete objects into a holistic framework. Digital humanist criticism of the aggregate recognizes to a degree how the principles of organization reflect an ideological stance and can guide and construct scholarly discourse. This becomes evident when we observe the self-consciousness of various projects which aim to digitize, and make accessible for both human and machine readers, collections of texts. This section will consider the OCR projects of the *Early Modern OCR Project* and the *Improving Access to Text* project.

Optical character recognition, or OCR, represents the latest in a series of technological developments employed in the preservation, extension of accessibility, and, most importantly, the transmission of textual objects. A particularly robust definition of the process is defined by the IMPACT (Improving Access to Text) project, based in the EU:

> Optical Character Recognition (OCR) is the electronic translation of text-based images into machine-readable and editable text. Usually performed by software devices as part of a digitisation workflow, OCR works by performing a layout analysis of a digital image and breaking that image into smaller structural components to find zones of textual content. These zones will include the overall area of the page that features text. Within that zone the OCR software will identify individual lines of text, and within those lines will identify individual words and characters.
>
> Many OCR software suites are available for many types of use, and each runs to slightly different standards and methodology. At its simplest, however, all OCR software follows the same basic principle: once the software engine has identified a single character, it runs that character's properties through an internal classification of text fonts to find a match. It repeats the process for all characters within a word, and then runs that information through a dictionary of complete words to find a match. It extends this process through sentences, lines and text blocks until – ideally – all text in the image has been identified."[61]

At its core, OCR represents a process of turning the image of a page (or other discrete material unit of a text) into a readable text file. For humanistic endeavors, including the IMPACT project, but also others such as Texas A&M's *Early Modern OCR Project* (eMOP),[62] Leipzig University's

61. "IMPACT Strategic FAQ - Answers," IMPACT: Improving Access to Text, accessed January 27, 2017, `http://www.impact-project.eu/faqs/impact-strategic-faq-answers/#c517`.

62. "eMOP — eMOP," accessed January 27, 2017, `http://emop.tamu.edu/`.

Open Greek and Latin Project,[63] or the Library of Congress' *Chronicling America*[64] project for early American newspapers, OCR is wrapped in an ethos that exhorts ideals of access for research and teaching, and postulates an improved quality of life (of the mind) for scholars. This is obvious in the extension of the IMPACT acronym itself: Improving *Access* for Texts. And while this is understood as the ethics for why OCR-based projects and their outputs (i.e. OCR'd texts) are necessary and a force for good in the scholarly world, I use the IMPACT definition because of its perceived pragmatism in explaining the process while at the same time echoing traditional textual scholarship discussions about the nature of text.

Before I explain that statement further, however, it may be important to contextualize OCR and its specific place in textual discussions. There are new things OCR offers to bibliographical discourse, which will be explored, but OCR is also only a new technology in a series of technologies that have been used for the reproduction of text. Photographic reproduction, xerography, and microfilms have all been method or means that has been used to preserve, provide access to, and transmit images of pages since the late nineteenth century. These means of reproduction and transmission are familiar to those with extensive work in archives and libraries, as they are some of the primary ways ephemera and serial publications of previous centuries survived. Tanselle speaks in a detailed manner about these forms of reproduction, though primarily their uncertainty as witnesses to the text, though they may present a page image in such a manner as to be considered in absolute fidelity to the copy-text from which the image is derived.[65] Tanselle's constant refrain for criticizing modes of reproduction points to the "inherent uncertainty" of all modes of transmission.[66] According to Tanselle's rationale, photographic copies of images are not adequate replacements for the original printed texts; they are new documents themselves, as too many issues can affect the reproduction that would introduce variants to the work. Any form of page image, which excludes OCR due to the time of publishing, but is nonetheless relevant, can occlude documentary evidence that is noticeable only with the material text; the images can be manipulated, with emendations, annotations, or errors introduced by a human that would distinguish the photocopy from the printed copy-text (this issue, we can imagine, is exacerbated in a post-Photoshop world); and finally, the collection of page images that make up a whole text, may not have been in fact created from a

63. "Open Greek and Latin Project — Digital Humanities," accessed January 27, 2017, `http://www.dh.uni-leipzig.de/wo/projects/open-greek-and-latin-project/`.

64. "Chronicling America: Historic American Newspapers," accessed January 27, 2017, `http://chroniclingamerica.loc.gov/`.

65. Tanselle's exact definition of reproduction understood here is: "the product of any chemical or electrostatic process that aims to represent with exactness (though perhaps on an enlarged or diminished scale) not only the text of a given document but also the details of its presentation, insofar as they can be duplicated on a different surface." G. Thomas Tanselle, "Reproductions and Scholarship," *Studies in Bibliography* 42 (1989). 26, ISSN: 0081-7600, accessed January 1, 2017, `http://www.jstor.org.ezproxy.library.tamu.edu/stable/40371897`

66. Thomas Tanselle, *A Rationale of Textual Criticism* (University of Pennsylvania Press, 2011). 43, ISBN: 978-0-8122-0042-3, accessed February 1, 2017, `http://muse.jhu.edu.ezproxy.library.tamu.edu/book/3599`.

single copy-text at all, but from multiples, should scans of pages from other copies be necessary in cases of damage or other effects that could obstruct reading. It is for all these reasons that Tanselle complains of any trust one would put into a reproduction of a text whilst believing it to be a faithful reproduction that unerringly transmits the original printed source to a photocopy, or more relevant here, a digital image.[67]

Tanselle illustrates a necessary divide between ways to approach a reproduction of a text–that between a belief that a reproduction can be a replacement, or even a viable stand-in for the printed text it is derived from, and the acceptance that textual reproductions are a new incarnation of the text that has sprung into existence, sans any authorial intent, with its own intricacies and forms of documentary evidence that bear upon the work. For Tanselle's reasons it's to admonish bibliographers who, to him, should know better than to entertain ideas that replacements for print editions of texts could be exist.[68] Tanselle's insistence that the facsimile reproductions of texts are new documents in their own rights is empowering for the processes that reproduce texts; Tanselle's line of thinking allows us to approach OCR'd versions of texts (as technological successors of textual reproduction) within the bounds of bibliographical theory insofar as we become able to claim an OCR'd reproduction of a text as a new edition of a text, designed and implemented as an editorial method of textual transmission in the creation of a new text for reading a work.[69] Tanselle's fears should not be understood as only focused on visual reproduction, however. The inherent uncertainty of a text is brought up again in the context of OCR, photocopy, etc. because this is their primary means of addressing textual information, but even more traditional bibliographic pursuits approach this issue. Enumerative bibliographies such as Wright's similarly, in their descriptions, must recognize that they only ever approach a concept of the text at a certain level that is not necessarily as microscopic as the ideal would prefer. That ideal being the observance, comparison, and collation of every copy of an edition, and every edition of a work. OCR projects and bibliographies alike must constrain themselves in the interest of practice, but in doing so assume biases and

67. Tanselle's work cited here is also a robust literature review and reading list of sources for the history of photographic reproduction. Though dated, the article itself can provide a thorough account of much early scholarship and discussions of these forms of technological textual reproduction. For Tanselle, however, many of the pieces are flawed because of their lack of discussion of the limitations and the overeagerness of scholars to discard their incredulity for these textual artifacts and to accept them as replacements for the originals. Tanselle, "Reproductions and Scholarship"

68. Ibid. 26.

69. Tanselle is not alone in his thoughts. What Tanselle labels uncertainty about a text, Peter Shillingsburg has acknowledged the premise that part of accepting the nature of a text–whether or a book or a manuscript–is to accept that one instance of the text cannot exist in the same physical space as another instance of text. By extension, Matthew Kirschenbaum has acknowledged the same by extending the incongruence of textual objects to the digital realm; while maintaining their still material nature, Kirschenbaum argues that digital access to texts recalls unique physical operations on the part of the mechanism being used at the moment of access. Peter L. Shillingsburg, *From Gutenberg to Google : electronic representations of literary texts. Peter L. Shillingsburg* (Cambridge, UK ; New York : Cambridge University Press, 2006., 2006). 13, ISBN: 978-0-521-86498-5; Matthew G. Kirschenbaum, *Mechanisms : new media and the forensic imagination. Matthew G. Kirschenbaum* (Cambridge, Mass. : MIT Press, [2008], 2008), ISBN: 978-0-262-11311-3

judgments (often in favor of first editions) of which versions of a text or work to privilege in their description and representation.

Returning to the IMPACT statement, and its relevance to bibliography, the statement calls attention to the primary purposes of OCR: the creation of ”machine-readable” and ”editable” text. The output of OCR is defined primarily by its affordance to be interpreted and manipulated by a machine. The word *editable* is important here because it implies that there is a necessity in the output of OCR to be changed, that it will need a human hand to intervene between the process of the OCR software reading an image, and the output being used for textual analysis, or whatever other function it may be accessed for. The primary audience, as related by its requirement to be ”machine-readable,” is that of the computer, and not the human reader. The human reader, of course, has the original .pdf or page image that they can always read themselves. I would additionally argue the Boolean logic of *and* twists the two concepts of machine-readability and editing into co-dependent objects. Machine-readability is necessarily tied to data's ability to be reconfigured and modeled, either because the machine will read the text in ways that naturally deform it, or the methods of studying the text computationally as determined by the human scholar will demand non-standardized modes of reading information.[70]

Following the IMPACT statement to the second paragraph, the way this is accomplished is, interestingly, resonant with early modern methods of setting type, that is, with the focus of both OCR and hand-press methods of printing, where the individual letter is a discrete unit of the composition process. I compare this against the nineteenth-century methods of stereotype and electrotype printing, which render the printing process to the level of the page. Textual criticism and analytical bibliography have made the sites of their study in the hand press printing period the moment where human labor or decision-making has an effect on the production of the text; most commonly, this seems to revolve around the introduction and transmission of errors, but otherwise any sort of variant, whether or not they are labeled so harshly as an error, among editions or copies of texts will certainly fall under the scrutiny of the material disciplines. The importance of human labor and decision-making, however, is what I would like to focus on in regards to OCR, however. Bibliographical views on texts acknowledge that errors occur not because of faults in the machinery that printed the text, but because of mistakes made on the part of a human agent that is located at some point between and inclusive of the author and the reader.[71]

70. Stephen Ramsay argues this idea in *Reading Machines*, wherein the possibility of computation interpretation for Ramsay is wholly determined by one's ability to do what is not possible in printed, material text: word frequencies, spatial or semantic changes to texts, and applicability of algorithms become the default premises of what digital hermeneutics is all about. He provides a glancing mention of textual scholarship and analytical bibliography in this discussion by foregrounding what he sees as a shared practice of ”expos[ing] the bare facts of a text” between algorithmic criticism and the material disciplines, calling such work a ”prelude” to the critical methods that follow this fact-determination process. Stephen Ramsay, *Reading Machines: Toward an Algorithmic Criticism* (Urbana, IL: University of Illinois Press, 2011). 6

71. This sentiment is rooted in D. F. McKenzie's *Bibliography and the Sociology of Texts*, wherein his use of the

What needs to be explained more thoroughly, then, is the way in which a process of transmission such as OCR does involve the same capabilities for one to make judgments, interpretations, and decisions about the object of the OCR process. In this instance, it is worth looking at the Early Modern OCR Project (eMOP), based out of Texas A&M's Initiative for Digital Humanities, Media, and Culture. eMOP can be understood in this discussion, to be a way for early modern texts to become accessible to scholars beyond the ways they are traditionally. Like the IMPACT project's statement, the mission statement of eMOP can be read in one way as reinforcing the idea of what OCR is by means of the purpose of its output, but, as well, their mission statement introduces a sentiment more akin to the editorial statements of critically edited text.[72] That is, their mission statement presents a scholarly viewpoint that guided their process, or, in other words, their guiding principles:

> The Early Modern OCR Project (Lead PI, Dr. Laura Mandell) is an effort, on the one hand, to make access to texts more transparent and, on the other, to preserve a literary cultural heritage. The printing process in the hand-press period (roughly 1475-1800), while systematized to a certain extent, nonetheless produced texts with fluctuating

word *sociology* as it relates to bibliography hinges upon the recognition of the human element in the creation of texts; McKenzie relies on this in order to ensure that his description of sociology does not become as understood as rigidly empirical as the tradition of bibliography (that is, primarily, descriptive) he is trying to supplant. The human agency embedded in McKenzie's sociology of the text, then, is responsible for its ability to tap into "social motives":

> By dealing with the facts of transmission and the material evidence of reception, it can make discoveries as distinct from inventing meanings. In focussing on the primary object, the text as a recorded form, it defines our common point of departure for any historical or critical enterprise. By abandoning the notion of degressive bibliography and recording all subsequent versions, bibliography, simply by its own comprehensive logic, its indiscriminate inclusiveness, testifies to the fact that new readers of course make new texts, and that their new meanings are a function of their new forms. The claim then is no longer for their truth as one might seek to define that by an authorial intention, but for their testimony as defined by their historical use."

D. F. McKenzie, *Bibliography and the sociology of texts. D.F. McKenzie* (Cambridge : Cambridge University Press, 1999., 1999). 15,29, ISBN: 978-0-521-64258-3, accessed February 1, 2017, `http://lib-ezproxy.tamu.edu:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=cat03318a&AN=tamug.1868498&site=eds-live`

72. The MLA provides a standard in its guidelines for scholarly editions that discusses these sorts of statements:

> Scholarly editions generally include a statement, or series of statements, setting forth the history of the text and its physical forms, explaining how the edition has been constructed or represented, giving the rationale for decisions concerning construction and representation. This statement also typically describes or reports the authoritative or significant texts and discusses the verbal composition of the text—its punctuation, capitalization, and spelling—as well as, where appropriate, the layout, graphic elements, and physical appearance of the source material. Statements concerning the history and composition of the text often take the form of a single textual essay, but it is also possible to present this information in a more distributed manner.

"Guidelines for Editors of Scholarly Editions," Modern Language Association, accessed February 11, 2017, `https://www.mla.org/Resources/Research/Surveys-Reports-and-Other-Documents/Publishing-and-Scholarship/Reports-from-the-MLA-Committee-on-Scholarly-Editions/Guidelines-for-Editors-of-Scholarly-Editions`

> baselines, mixed fonts, and varied concentrations of ink (among many other variables). Combining these factors with the poor quality of the images in which many of these books have been preserved (in EEBO and, to a lesser extent, ECCO), creates a problem for Optical Character Recognition (OCR) software that is trying to translate the images of these pages into archiveable, mineable texts. By using innovative applications of OCR technology and crowd-sourced corrections, eMOP will solve this OCR problem.[73]

This statement contextualizes the collection of texts, largely derived from the Early English Books Online (EEBO) and Eighteenth-Century Collections Online (ECCO) databases, with particular bibliographic features that had to be approached as conscious issues inherent to texts from the hand-press period. Just as with the IMPACT statement, eMOP presents a conscious consideration for the outputs of the OCR process, but where the two differ is in the acknowledged purpose of the process, and their individual projects' thoughts as to the scale of their work. In comparison to the IMPACT statement, which refers primarily to the individual text, and the individual components that make up a document in the OCR process, the eMOP project's approach to early modern texts considers a more broad view; an individual text is displaced in favor of the aggregate that makes up a "literary cultural heritage." The stated purposes of the outputs of the OCR'd texts corresponds to this larger view, wherein the OCR'd texts can be considered viable because they are "archiveable" and "mineable." Both terms specifically place the texts within a context that defines the individual artifacts by their position within a larger collection; first, the term *archiveable* declares a necessity for an overarching structure that surrounds the items within the collection. Archives, are, in the minds of digital humanists and literary critics, collections of objects that are organized by a taxonomy that is both informed by and informs litery scholarship as to the relationship amongst the materials, rather than their individual values.[74] In more direct words, texts found in the EEBO and ECCO collections are defined by their periodicity–they belong to a category that is innately literary, as described by Mandell.[75] The important shift that occurs here,

73. "eMOP — eMOP."

74. Kate Theimer describes the discrepancy between the digital humanist approach to archives and that of formally trained archivists. For Theimer the divide between conceptions of the archives hinges on matters of selection and authenticity. In my statement above, I account for the idea of selection, and how materials in a collection are organized that seems to scholars the most important detail of the collection. Theimer, however, notes that archivists are not in the business of selection, but instead in contextualization. Following Tanselle's belief that reproductions of a text are new documents and not replacements for the original text, we can understand the selection-based idea of archives as more fitting for materials that are not, in Tanselle's sense, part of any historical context that would give archivists cause to associate them with the original materials.Kate Theimer, "Archives in Context and as Context," Journal of Digital Humanities, June 26, 2012, accessed February 11, 2017, `http://journalofdigitalhumanities.org/1-2/archives-in-context-and-as-context-by-kate-theimer/`

75. Mandell points to early publications of poetry anthologies that organize literature according to historical and canonical interest and the idea of the "Romantic" period emerging at the same moment as the poetry it refers to.Laura Mandell, "Digitizing the Archive: The Necessity of an "Early Modern" Period," *Journal for Early Modern Cultural*

though, and which is reflected in eMOP's ideological approach to OCR, is the idea that it is not the aesthetics that should inform periodization, but rather the medium, the history of the book and its production, that should define the means of collection and presentation, and in the case of eMOP, the actual methods that undergird the OCR process.[76]

The term *mineable* carries the same connotation. The mining of texts (which refers to the text mining, or computational analysis of and algorithmic operations run on a collection of texts) also necessitates an aggregate. In cases in which smaller collections or individual texts may be concerned, what can be deemed "the aggregate" are the words, however, the most opportune uses of text mining, or "distant reading", look to collections comprised of whole documents. This returns to the purpose of OCR's outputs as understood by those involved with their production: if one of the primarily understood reasons for why a text will be used is that it's individual features will be enmeshed within the dominant patterns visible at scale, then the process to edit and represent these texts, must take in mind the dominant patterns of textual production. This is primarily Mandell's point in reinforcing the idea of the early modern period in the digital age via the means of production of texts over semantic or historical categories that have also defined periods:

> "...whether thinking about the difference between coterie print and mass print cultures, the or taking account of our current encounter with 'big data' and 'distant reading,' shoving the eighteenth century out of modernity and into early modernity suggests even more: that *scale* matters most about medium as a system of production, circulation, and dissemination."[77]

eMOP's statement puts Mandell's claim into a practice by primarily describing the objects of their work confined to the humanistic early modern period as coterminous with the bibliographic hand-press printing period, and by addressing the details this period in terms of its mechanical flaws: "fluctuating baselines, mixed fonts, and varied concentrations of ink (among many other variables." It is only through understanding these innately bibliographical processes that the texts can be prepared for future use. What Mandell and the eMOP statement both understand is that the semantic, aesthetic, and theoretical analysis of literature is predicated entirely on the material conditions of the text, which forms the foundation of its ability to be transmitted. This is a point bibliography understands well, but where OCR expands this idea is in its capacity to make the hand-press period a scholarly literary period in itself, as the mechanical process that created the texts in the collection become the overall category that defines the collection which will become subject to scholarly

Studies 13, no. 2 (March 8, 2013). 83, ISSN: 1553-3786, accessed February 11, 2017, doi:10.1353/jem.2013.0019, https://muse.jhu.edu/article/501759

76. Ibid. 90.

77. ibid. Empasis Mandell's

questions of political, aesthetic, and otherwise semantic exploration of language that may be considered by the scholars asking the questions to be independent from the processes of production.

For bibliography and the material disciplines to adequately study the transmission of texts, they must account for the concept of scale as a capacity for texts to be amongst one another in a collection. Enumerative bibliographies have possessed to some degree an awareness of this concept, as they have made their *modus operandi* the arrangement and presentation of texts, though not without issues. Where the digital humanities can make an intervention is in improving how bibliographic collections of texts can be conceived and in continuing to question how it is their data and the assumptions and interpretations that undergird the construction of that data become representative of ideologies and biases within a dataset. If we are to understand that both scholars creating digital collections and enumerative bibliographers serve as guides and facilitators of access to textual history, from which cultural scholarship is able to progress, then beginning to question the processes that inform aggregation and the accumulation of material for presentation becomes an important part for the research process. It is in recognizing that how we are able to or allowed to see the texts that determines how we read them.