

# The Sociable Textual Archive: Laying the Groundwork for Linked Bibliographic Entities

Brent Nelson

Volume 3, numéro 1, 2019

Methodologies of Understanding and Enacting Open Scholarship

URI : <https://id.erudit.org/iderudit/1084222ar>

DOI : <https://doi.org/10.5334/kula.12>

[Aller au sommaire du numéro](#)

Éditeur(s)

University of Victoria Libraries

ISSN

2398-4112 (numérique)

[Découvrir la revue](#)

Citer cet article

Nelson, B. (2019). The Sociable Textual Archive: Laying the Groundwork for Linked Bibliographic Entities. *KULA*, 3(1), 1–11. <https://doi.org/10.5334/kula.12>

Résumé de l'article

Much of our scholarly thinking of the 'social' in digital editing has been with respect to the human processes of building an archive or an edition. This paper explores the idea of the 'social' with respect to the archive's materials themselves. Taking the John Donne Society's Digital Prose archive as a test case, this paper explores the infrastructure and resources available for creating an open 'sociable' knowledge network of linked bibliographic resources.

Copyright © Brent Nelson, 2019



Ce document est protégé par la loi sur le droit d'auteur. L'utilisation des services d'Érudit (y compris la reproduction) est assujettie à sa politique d'utilisation que vous pouvez consulter en ligne.

<https://apropos.erudit.org/fr/usagers/politique-dutilisation/>

**é**rudit

Cet article est diffusé et préservé par Érudit.

Érudit est un consortium interuniversitaire sans but lucratif composé de l'Université de Montréal, l'Université Laval et l'Université du Québec à Montréal. Il a pour mission la promotion et la valorisation de la recherche.

<https://www.erudit.org/fr/>

---

## METHOD

# The Sociable Textual Archive: Laying the Groundwork for Linked Bibliographic Entities

Brent Nelson

University of Saskatchewan, CA  
[brent.nelson@usask.ca](mailto:brent.nelson@usask.ca)

---

Much of our scholarly thinking of the ‘social’ in digital editing has been with respect to the human processes of building an archive or an edition. This paper explores the idea of the ‘social’ with respect to the archive’s materials themselves. Taking the John Donne Society’s Digital Prose archive as a test case, this paper explores the infrastructure and resources available for creating an open ‘sociable’ knowledge network of linked bibliographic resources.

---

**Keywords:** FRBR; digital archive; literature; bibliography; John Donne

---

## Introduction

Much of our scholarly thinking of the ‘social’ in digital editing has been with respect to the human processes of building a digital archive or an edition (Siemens et al. 2012; Robinson 2015 and 2016) or an attempt to reflect the complex social relationships involving various human agents, institutions, processes, and textual manifestations leading up to and proceeding from a moment of publication, in what ever form that might take (McGann 1992; McKenzie 1999). This paper explores the idea of the ‘social’ with respect to the archive’s materials themselves. In the age of Web 2.0 and Linked Open Data (LOD), we at some level expect our materials to be able to play nicely together. In what follows, in recognition of this shift in focus, I use a slightly different version of the term: ‘sociable.’ The added element here is the emphasis on potential, the *openness* of linkable data. At any given moment, an entity might be social, that is, involved in some sort of relational nexus, but to be *sociable* is to be inclined or open to involvement in a given social context. One might be social but not open: one might be involved within an already established and determined set of relationships, but in this context, to be sociable is to be prepared and ready for new social involvement.

In this paper, I will begin to lay the groundwork for considering the possibilities of Linked Open Data for bibliographic entities in the context of a particular literary archive: the John Donne Society’s Digital Prose Project, and ultimately, a complete Donne Archive that would include the Donne Variorum Project’s Digital Donne. Similar attempts at LOD with respect to person entities are underway at various points of intersection between scholars, librarians, and digital humanities practitioners. Central to these considerations are various forms and expressions of authority control expressed by authority files, such as the Library of Congress Authority, or the aggregating service of VIAF. In a previously published paper on the possibilities of LOD for modeling early modern social networks, I posit that the possibility of referencing these authorities in the context of individual projects takes us a long way to making diverse and disparate datasets sociable, that is, open to relationships with other datasets on the basis of named and identified person entities who already have some place in the official historical record (Nelson 2014). There are many challenges in this area—namely, what to do with historical persons who are not already part of the official record and therefore not included in authorities—but basic infrastructure is in place that could be refined and developed in new directions to enable any variety of digital projects with overlapping social and historical contexts to point from (for example) tagged entities in TEI documents to their authority headings in one of the available authorities on the Web.

The case of bibliographic entities, however, is much more complicated than that of personal entities. The key step in a sociable archive, that is, one that is open to bibliographic relationships, is some sort of authority

or standard for unambiguously naming and referencing entities. There is, however, no such standard for bibliographic entities at the top-level of the 'work.' In this domain, the needs and expectations of scholars depart rather sharply from those of librarians and information specialists. An example from VIAF will serve to illustrate. The person entity for John Donne, the seventeenth-century poet, is clear and unambiguous. There is one, clear heading. The case for Donne's works, in contrast, is very problematic. To begin, there is no clear and consistent distinction in VIAF between a 'work' and an 'edition.' There is a search function for 'work,' which is able to return an unambiguous result for Margaret Atwood's *A Handmaid's Tale*, for example; but a search for Donne's *Biathanatos* returns two entries: one with the heading 'Donne, John, 1572–1631. |Biathanatos,' and another with heading 'Sullivan, Ernest W., 2. |Biathanatos,' which, from a scholar's point of view, is not a 'work' but rather Sullivan's edition of Donne's *Biathanatos*. The full listing of 'works' under Donne's personal entry is rife with such problems. There is a heading for 'Sermons. Selections,' which is clearly not a 'work,' and there is not a single entry for an individual sermon, which is what a literature scholar *would* call a 'work.' For the purposes of a digital archive of Donne's prose, the former is of little value; it is the latter that we need to be able to point to, the individual sermon. The case of Donne's lyrics is similar, but with exceptions. There is an entry in VIAF for his lyric poem 'The Flea' as a 'work,' but only a few of the dozens of lyrics in Donne's oeuvre are represented in this way—though, again, the 'Songs and Sonnets' as an (imprecise) generic grouping are identified with a separate heading as a 'work.' So then, 'genres,' 'editions,' and 'works' are all confusedly identified as 'works,' and the list of 'works' proper is radically incomplete. There is no value in VIAF for an implementation of LOD in the context of a literary archive, and the Library of Congress has no authority for 'work' at all.<sup>1</sup> From the outset, then, we are at a disadvantage in dealing with bibliographic entities because the same institutions that have developed a key piece of infrastructure for person-based LOD present us with no such infrastructure for their core domain of concern: bibliography.

What would it take to develop a central authority for 'works' similar to the available authorities for people? To frame this exploration, it makes good sense to adopt as a heuristic starting point the Functional Requirements for Bibliographic Records (FRBR), a popular high-level set of recommendations that is very popular in information science circles in theory if not yet in practice.<sup>2</sup> FRBR is nonetheless a helpful tool for thinking about how to structure bibliographic resources, but as we will see, there are some complications in our example that are not easily resolved within a FRBR framework. The FRBR model for describing bibliographic resources posits four types of entities for identifying 'products of intellectual or artistic endeavor (e.g. publications)': 'work,' 'expression,' 'manifestation,' and 'item' (IFLA 1998). The OCLC Research website (2017) summarizes these entities as follows:

- *the work*, a distinct intellectual or artistic creation
- *the expression*, the intellectual or artistic realization of a work
- *the manifestation*, the physical embodiment of an expression of a work
- *the item*, a single exemplar of a manifestation.

### The 'Work' and the 'Expression'

From a broadly theoretical perspective, the notion of the 'work' is complex enough to warrant a monograph (Smiraglia 2001). In the FRBR context, there are some conceptual and practical difficulties posed by both of the first two entities. On the surface, most readers can intuit the 'work.' In Maxwell's words, when we speak of a work, 'we are not thinking of a particular performance or publication of the work but of the intellectual creation that lies behind the various expressions of a work' (Maxwell 2008, 16; see also Coyle 2016, 3–28). A theoretical problem arises in the fact that a 'work' is an abstraction known only by one of its expressions manifest in concrete form, and these expressions and manifestations tend to vary—particularly for a writer who is defined by manuscript circulation, as was Donne (Marotti 1986; Stringer 2011). Is the altered thing then a new work? At what point does a variation in the form of expression result in a new work? Are Milton's *Paradise Lost* in ten books and his *Paradise Lost* in twelve books distinct works or simply repackaged versions of the same work? One might argue that, from the perspective of typology, a biblical epic in twelve books really is quite a different thing than the same poem in ten books. To push the issue further, is Nahum

<sup>1</sup> Stahmer (2016, 26) identifies an OCL namespace (<http://experiment.worldcat.org/entity/work>) for works, but at the time of writing this article, this url is dead. He cites also a Library of Congress namespace (<http://id.loc.gov/authorities/names>), but I see no evidence of an implementation of an authority mechanism for 'works.'

<sup>2</sup> Implementation of these recommendations has been slow. For discussion and criticism of some of these implementations see Welsh (2017) and Coyle (2016) as recent examples. I wish to thank my one very helpful reviewer who supplied generous and very helpful recommendations, including these citations.

Tate's happy-ending *King Lear* the same work as the first-folio *King Lear*? There would be a few quibbles among Donne scholars on the question of what constitutes a new work. FRBR would suggest that Donne's *Conclave Ignati* and his English translation of it, *Ignatius His Conclave*, are a single work, though Donne scholars might cavil at a simple identification. And yet, Geoffrey Keynes's authoritative *Bibliography of Dr. John Donne* (1973) treats them as a single work (19). In most instances, the academic question of precisely where the line should be drawn between an expression and a whole new work can, as in Keynes's case, be resolved easily enough for the sake of practicality.

Nonetheless, even in practical terms, the identification of a work is considerably more complicated than the identification of a personal entity, where a biological being is (despite complications of hybridity and identity) almost always an easily understood referent. In early modern printing, titles are often complex and therefore shortened in common usage and can vary across editions, although these can similarly be resolved easily enough with a standard short title articulated in a one-to-many relationship with all title variants. The case of Donne is complicated, though, by several works for which there is no authorial title (and therefore many surmised titles), such as many of Donne's poems and many of his sermons. There are also practical problems—and this is crucial to the case being examined here—in the way our bibliographic infrastructure (chiefly, but not only libraries) handles the 'work.' Libraries do not catalogue works, but only manifestations of works (specifically, in this case, editions). So, a library will catalogue Donne's *LXXX Sermons*, a title which refers not to any particular work, but rather to many (i.e. eighty), unnamed works. In fact, this case is even more complicated than that, because this book also contains Izaak Walton's life of Donne, which is also a work in its own right. Library catalogues seldom if ever record the works, that is, the individual sermons in this case. This is a practical problem, but it points also to a theoretical problem. While it seems obvious to a Donne scholar that the sermon is the work and this edition is a collection of works, can the same be said of Donne's paradoxes and problems, which circulated as a group? And while it makes no scholarly sense to think of the 'Songs and Sonnets' as a 'work' as VIAF apparently does (these poems circulated individually, not as a clearly defined group), should Donne's *Satyres* be considered a work, given that they always circulated as a group of five, and therefore seem together to constitute an artistic unity? In any case, serious thought needs to be given to the way in which this relationship between work and edition is expressed in such cases of a many-to-one relationship.

### The 'Expression' and 'Manifestation'

In the context of Donne, an 'expression,' understood as 'the intellectual or artistic realization of a *work* in the form of alpha-numeric, musical, or choreographic notation, sound, image, object, movement, etc.,' points to a helpful distinction between a sermon as oral product and the same sermon as an inscribed object (IFLA 1998, 19). The oral sermon might then have any number of manifestations, i.e. preachings, although there is no indication that Donne ever preached the same sermon twice. A Donne scholar could quite comfortably understand the oral and the written as two expressions of a single work, although the language of FRBR might raise the question of whether these are not in fact two distinct works. As much as a printed sermon might in a single edition vary from copy to copy, the possibilities for intervention and therefore alteration in an oral context are significantly greater. Rarely, in any era, would a sermon preached on two different occasions not vary significantly: vagaries of memory and circumstance would ensure divergence. Then there is the question of the notes a congregant might take on a sermon that was preached. Sometimes the notes might be quite extensive, but even so, would this be an expression or a new work? So, to summarize: in the FRBR framework, a 'work' might be Donne's sermon on Psalm 144:15, which in one 'expression' was an oral discourse delivered as a performance (a 'manifestation') on April 20, 1620, at Whitehall, and in another expression was an inscribed version that ended up in several 'manifestations,' including an original holograph in manuscript (now lost), a printed artefact contained in *LXXX Sermons* (1640), and several other manifestations in manuscript.

### The 'Item'

One might further say that, although libraries catalogue editions, they are in fact primarily, sometimes almost exclusively, interested in the *item*: the particular copy of an edition, that is, one that they own and therefore want to keep track of. From the perspective of a Renaissance scholar, the item is important for somewhat different reasons. Every printed book from the early days of print is unique even within a particular edition: there are almost always stop-press variants throughout any given edition, which means that any two copies might very well have slightly different type-settings. Moreover, in the history of a book, various interventions occur that are of potential interest to scholars: bindings, marks of ownership, annotations,

and so on. Scholars would want to know, for example, that the copy of Donne's *LXXX Sermons* in the Geisel Library at University of California at San Diego was owned by Professor Don Cameron Allen and bears his annotations and that it is bound with Donne's *Fifty sermons* (1649), also with his annotations. But with the 'item' in the FRBR framework we run into problems of nomenclature: as noted above, we want to be able to point to and name manifestations of individual works (sermons in our example) that occur (confusingly) within manifestations that we know as editions (collections of sermons). Following the naming practices of the Text Encoding Initiative (TEI, the XML standard for mark-up of printed books and manuscripts), it would be natural to call these things 'items,' but in a FRBR framework, this term is already in use for a different purpose, as articulated above.

As implied throughout the above discussion, the crucial attribute and functional requirement for all of these entities is the 'name.' For a resource to be sociable, it needs to have a reliably and commonly understood name. The key element of infrastructure for sociability among bibliographic resources, then, is the Uniform Resource Name (URN), i.e. a name that can serve as a unique identifier for a resource. These might take the form of a scholarly siglum (a convention used in scholarly editions) or any unique identifier attached to a bibliographic entry in a standard bibliographic source. Once a bibliographic entity is named, it can be addressed, and therefore has the potential to become sociable. The naming mechanism for linking these entities on the Web (the Universal Resource Identifier) is the infrastructural implementation of the URN, and beyond the scope of this paper: as noted, the first step is to sort out the question of naming (Wood et al 2014, 11; see this also for a good explanation of URIs).

So, then, library infrastructure does not supply an expressive and granular naming system for linked bibliographic resources, and to build one seems an impossibly complicated task—even for a small corner of the bibliographic world, such as English Literature. For these reasons, I have serious doubts about a large-scale LOD solution for bibliography; so I will begin at the opposite end of the scale, with just the Donne Digital Archive—and a sub-set of relevant bibliographic resources—to seek a model that might be scaled up for more general implementations. Despite the deficiencies in the resources I describe above, there are crucial bibliographical works available and under development in the library domain (the new English Short Title Catalogue at the British Library, for instance) that would be of potential value for an archive of Donne's works. There is also a significant body of scholar-produced bibliography related to Donne's literary corpus. These resources are carefully structured, well rationalized, and in some cases correlated to other standard reference works, such as the old English short title catalogues—Pollard and Redgrave (1475–1640) and Wing (1641–1700)—and more recently, Early English Books Online (EEBO). Further, recent editorial projects such as the Donne Variorum edition of the poems and the new Oxford University Press edition of Donne's sermons have developed a set of short forms for referencing many of Donne's works. Altogether, these resources could be helpful in developing infrastructure for distinguishing and relating works and particular manifestations of works relevant to Donne.

The rest of this paper will use the case of John Donne to explore the challenges and possibilities for using conventional identifiers in bibliographic resources (both primary texts and secondary materials) as URNs to facilitate internal linking but also for linking outward to other, institutional and commercial materials related to the works of John Donne. The John Donne Society's Digital Prose Project was initiated by the Society in 2009 with the idea of developing for Donne's prose the sort of digital archive that had been developed for Donne's poetry by the Variorum project, known as DigitalDonne (Stringer n.d.; Nelson 2013). From its inception, it was understood that the prose archive would be built by a community of interested and invested users (in this case the John Donne Society in the first instance—and its immediate network), but also any and all professional or citizen scholars who would like to participate (Nelson 2013, 196–7). As a volunteer-driven project, our goal is to complete phase one (which will include at least one XML-encoded transcription of each of Donne's prose works) without any major funding. By the end of phase one, we will have in hand a transcription of at least one and possibly more than one witness of each prose work, with matching images of the source document, as well as images of several other witnesses (both print and manuscript), and an assortment of other already digitized sources presented by the Donne Variorum. These, in addition to a number of other standard, structured bibliographic resources that are 'sociable' (in that they are easily addressed) make Donne a good case study for exploring the possibilities of linked data for developing an enriched reading environment. In what follows, I have altered the FRBR nomenclature to one more familiar to textual scholars: work; document; copy; item. Specifically, because TEI is the literature scholar's standard for representing texts, I have elected to use 'item' in the TEI sense as an item in a manuscript or printed book, rather than the FRBR sense of what we would see as the equivalent of an edition.



## The English Short Title Catalogue (ESTC) and Early English Books Online (EEBO)

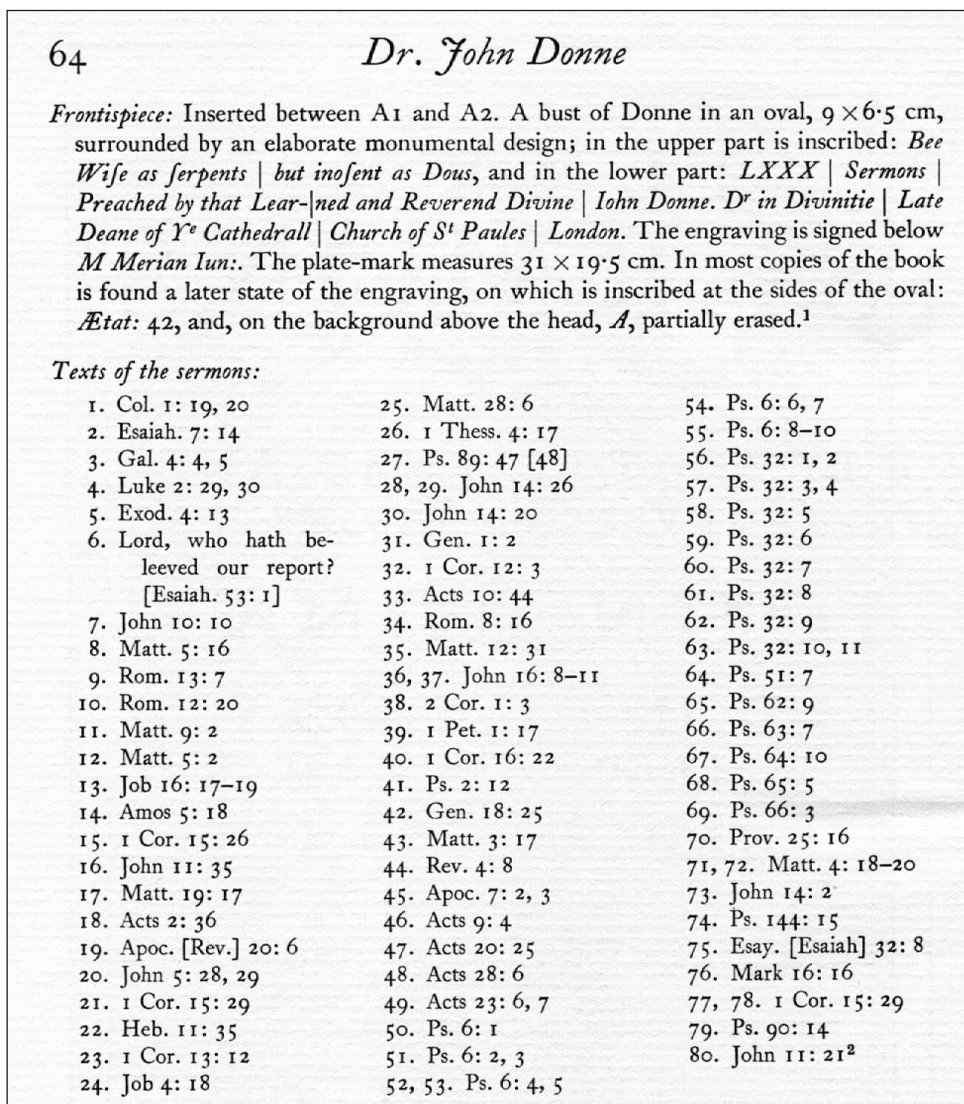
*The English Short Title Catalogue* (ESTC) effectively supersedes Pollard and Redgrave (1976; STC) and Wing and Pollard (1972).<sup>3</sup> It also includes, in some cases, references to *Early English Books 1475–1700* (EEBO) and *Eighteenth Century Collections Online* (ECCO), which also reference STC and Wing. For referencing purposes, the ESTC provides an ESTC citation number (in this case S121697) for each edition (manifestation), which is the unique identifier in their provided ‘Permalink’ (see **Figure 1**).

The EEBO entry for *LXXX Sermons* references back to the old STC (STC [2<sup>nd</sup> ed.], 7038) and also to Geoffrey Keynes’s specialized *Bibliography of Dr. John Donne* (Keynes, G. Donne [4th ed.], 29) as well as to the microfilm series Early English Books (Early English Books, 1475–1640, 12:1135), which is an obsolete form now that EEBO has superseded it and is accessible on-line. However, there is no short-form way to reference an EEBO entry (no reference number), even though there is a durable URL which contains a unique identifier. In the case of *LXXX Sermons*, EEBO does not supply a unique entry identifier of its own, but does provide STC, Keynes, and the microfilm as short-form identifiers of the edition it is describing (**Figure 2**).

**Figure 1:** English Short Title Catalogue (ESTC) entry for *LXXX Sermons*.

**Figure 2:** Early English Books Online (EEBO) entry for *LXXX Sermons*.

3 For a short, succinct comparison of how STC, Wing, the ESTC, and EEBO relate, see Gadd 2009, 683–4.



**Figure 3:** Keynes's entry for *LXXX Sermons* showing itemization of sermons (64).

More problematically for our purposes, what this durable URL actually points to is not the edition but rather one particular copy (and its images): there are several URLs in fact, pointing to several copies of this edition.

While the new ESTC provides an addressable unique identifier and points to other unique identifiers in sources it supersedes, it has no capacity for dealing with the works of Donne. Its only generic reference is to a 'Collective title'—'Sermons. Selected Sermons'—that in fact includes a number of works: namely, eighty individual sermons. By referencing Keynes, however, the ESTC associates with a reference work that does identify individual works, by means of the biblical text Donne preached on (**Figure 3**). So by this linkage, a near-universal resource for naming editions (ESTC) can leverage a niche resource (Keynes) to associate the edition with the works it contains. Although the ESTC does not account for works, it does point to a large number (though not exhaustively) of copies (which in FRBR are 'items'). If catalogued well, each of these copies, and more, should reference back to the ESTC, but more likely, they currently reference one of the short title catalogues it has superseded.

### **The Catalogue of English Literary Manuscripts 1450–1700 (CELM)**

In contrast to the ESTC, Peter Beal's *Catalogue of English Literary Manuscripts 1450–1700* (CELM) has a very different focus: its orientation is around work/manifestation, more specifically, the manifestation of a work as an item (not in the FRBR sense) in a manuscript. Under each author, the work is individuated, and then each manifestation of that work in a manuscript is assigned an identifier (author abbreviation + number). Poems, for example, are listed, and under each is a dedicated identifier for every instance of that poem in a manuscript. Likewise for each sermon (**Figure 4**).

**Sermon preached at Whitehall, April 30, 1620, on Psalms 144.15**  
 First published in *LXXX Sermons* (London, 1640), No. 74. Potter & Simpson, III, No. 2, pp. 73-90.

**DnJ 4024**  
 Copy in: the MS described under [DnJ 3998](#). c.1620s.  
 This MS collated in [Potter & Simpson](#)  
[Bodleian, MS Eng. th. c. 71, ff. 78-84v.](#)

**DnJ 4025**  
 Copy in: the MS described under [DnJ 4002](#). c.1620s-30s.  
 This MS collated in [Potter & Simpson](#).  
[Bodleian, MS Eng. th. e. 102, pp. 123-50.](#)

**DnJ 4026**  
 Copy in: the MS described under [DnJ 4004](#). 3 August-11 September 1624.  
 This MS collated in [Potter & Simpson](#).  
[National Library of Scotland, MS 5767, ff. 39r-44r.](#)

**DnJ 4027**  
 Copy in: the MS described under [DnJ 97](#). 1626-96.  
 This MS collated in [Potter & Simpson](#).  
[St Paul's Cathedral, MS 52. D. 14, Part I, ff. \[3r-32r\].](#)

Figure 4: Catalogue of English Literary Manuscripts 1450–1700 (CELM). <http://www.celm-ms.org.uk/authors/donnejohn.html>.

New York Public Library, Westmoreland Manuscript

Textual Apparatus

Copy-text: [NY3](#) Texts collated: B7 (f. 106v); B32 (f. 141); C2 (f. 15); C9 (f. 97v); CT1 (p. 232); DT1 (f. 120); H4 (f. 99); H6 (p. 32); IU2 (ll. 12–14 only, f. 50v); NY3 (f. [36v]); O20 (ff. 45v–46); SP1 (f. [60]); A (p. 38); B (p. 340); C (p. 340); D (p. 323); E (p. 323); F (p. 323); G (p. 320). 1633 Poems

Emendations of the copy-text: Heading: 16.] .-. Line 2 and] & 3 and] & 4 and] & 9 I love] llove and] & would be] wouldbe  
 11 that] y'

HISTORICAL COLLATION

Format:

Indentations: even numbered lines ind B7 CT1 DT1(plus 13) H4(plus 13);  
 ll. 2, 3, 6, 7, 10, 12 ind B32 C2 NY3 O20 SP1; ll. 2, 4, 6, 7, 10, 12  
 ind C9 H6; no ind A–G.

Miscellaneous: poem is last of four sonnets entered under the general HE Other

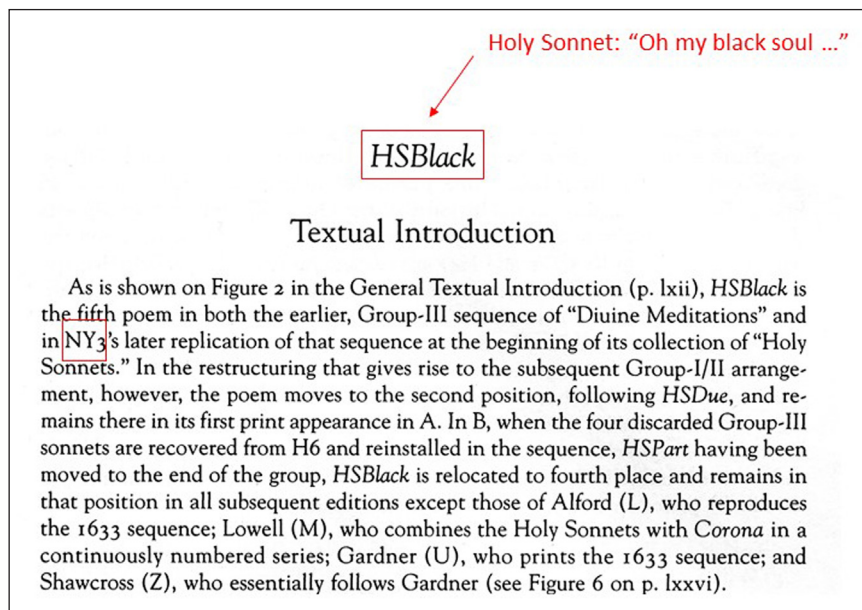
Figure 5: Textual Apparatus from Donne, *Variorum* (2005, 32).

However, while there is an identifier for each manifestation in a manuscript, there is no identifier for either the work or the manuscript (which is itself a manifestation, often containing many manifestations of works). Each of these is instead expressed in a long form. A further complication arises with Donne's paradoxes and problems, which are treated as a collective work, despite the fact that 'Paradoxes and Problems' is a generic category for what really amounts to a collection of individual works, much like each sermon is a work unto itself. Therefore, particular instances of any given paradox or problem has no siglum.

### The Donne Variorum

The Donne Variorum illustrates the value of short-form referencing systems in a complex knowledge environment (whether a siglum or some other unique identifier). It models an internal referencing system that is common enough in scholarly editions but particularly well executed in this instance. Given its purpose of disentangling a complex textual tradition that was born out of a body of literature that circulated widely first in manuscript and then in print, the Variorum introduced two key sets of short-form signifiers: a set of abbreviations for 'works' and a set of sigla for each manifestation of a print edition or manuscript. This enables a standardized, short-hand internal referencing system, as in this example of its textual apparatus (Figure 5).





**Figure 6:** Internal referencing in the Donne, *Variorum* (2005, 35).

Short forms are particularly useful in the context of early modern literature, where titles can often be quite long, are expressed in various forms, or sometimes lack a title entirely—as in the case of sonnets (as we have seen, a complication present in Donne's oeuvre). Thus there is added benefit in the uniformity of these sigla, again enabling a many to one relationship that resolves and reconciles different manifestations. For these reasons, short forms can also facilitate discursive discussion in cases where many sources are at play (**Figure 6**): The value of this system grows significantly when it is adopted outside the covers of the *Variorum*. The *Oxford Handbook of John Donne*, for example, uses these short forms when referencing Donne's works, as do volumes 3 and 4 of John Roberts's *John Donne: An Annotated Bibliography* (2004, 2013).

One significant limitation of the *Variorum* naming infrastructure, however, is a lack of granularity in naming manifestations: while it provides sigla for each manifestation in an edition or manuscript, it does not identify individual manifestations of works within these larger manifestations the way CEML does for manuscripts. One persistent problem in both FRBR and standard reference works for Donne is an inadequate accounting of the one-to-many relationship in manifestations at the collection level (commonly referred to as a 'manuscript') which in fact are composites of many manifestations of many works (e.g. individual poems in a miscellany).

### John Roberts's *John Donne: An Annotated Bibliography*

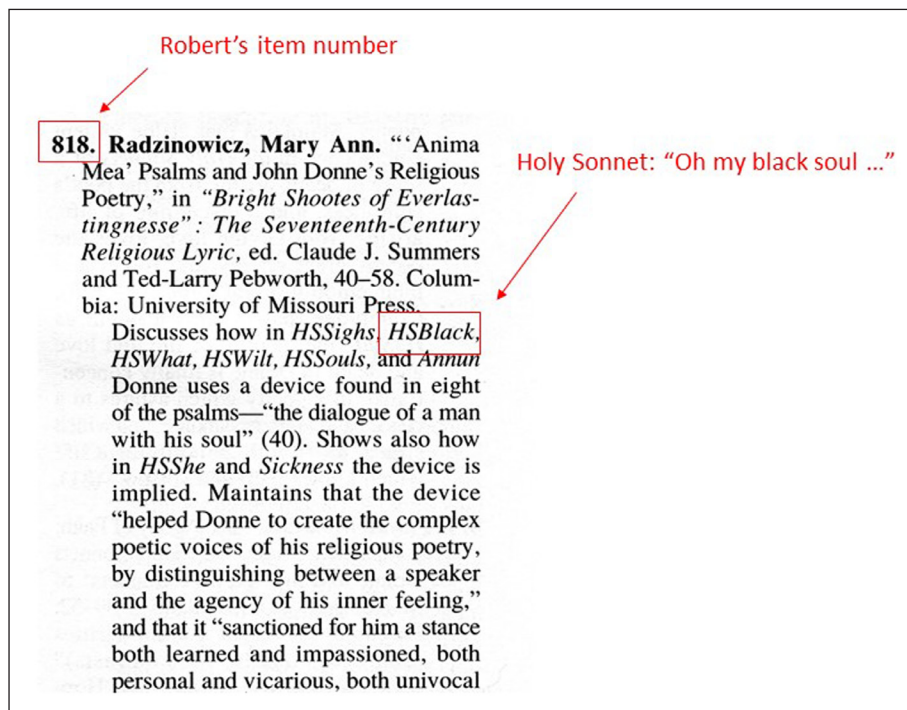
Roberts's four volumes of his *Bibliography* provide another sociable source with great potential (1973, 1982, 2004, 2013). In producing an exhaustive bibliography of all secondary literature on Donne from 1912 to the present, and assigning a number to each item, this resource in essence provides an authority for the secondary literature on Donne that can serve as a referencing system for correlating primary and secondary materials in our archive if we employ a convention of volume + entry number (which begins at '1' in each volume. See **Figure 7**).

So, the aim of the John Donne Society's Digital Prose Project is to model this sort of open reference-ability, pulling together a set of well-structured and standardized resources with sigla and other unique identifiers that can serve as URNs to form the basis of a sociable archive that links between resources and make its materials open to linking to and from resources outside the archive. To summarize:

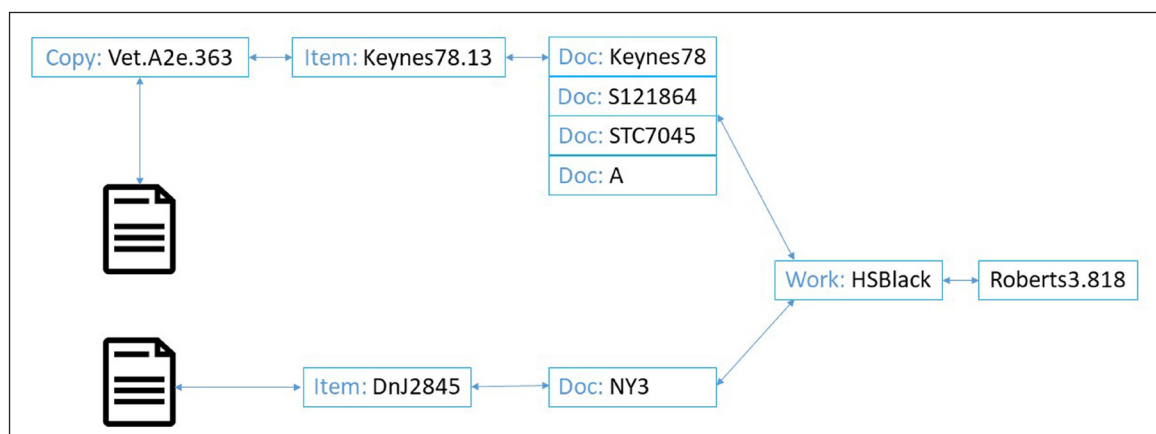
- Beal's *Catalogue of English Literary Manuscripts 1450–1700* (CELM) provides an identifier for each manifestation (item) in a manuscript; however, not all manifested works are itemized (e.g. paradoxes and problems), so new identifiers will need to be created for these, which undermines to some degree the value of CELM as a standard. Also, CELM does not employ a standard referencing model for 'work,' so we have to look elsewhere (the *Variorum*) for a naming standard for works. CELM also uses long-forms for manuscript names, but again, these can be mapped to the *Variorum* referencing system.
- The *English Short Title Catalogue* (ESTC) provides unique identifiers for each printed edition, with references to Keynes, the old STC, and Wing, but there are no such identifiers for manifestations of individual works; however, individual manifestations of each work can be supplied through linkage to Keynes, which itemizes the manifestations contained in each edition.

- The *Donne Variorum* provides short forms for each ‘work’ of poetry and some prose, but others will need to be developed for remaining prose works. It also provides sigla for manuscripts and print editions, but not for individual manifestations in manuscript and print: these can be supplied (mostly) by CELM.
- Roberts’ bibliographies provide unique identifiers for items of secondary literature, and in the final two volumes in the series, include references to work by means of the Variorum short forms.

So, together, these resources supply most of the naming infrastructure required for linking resources in a Digital Donne Archive. Note that what is imagined here is not an exhaustive FRBR-style accounting of all the materials involved, but only a means for naming and therefore creating links between materials, whether originals or digital surrogates. This, **Figure 8**, is how a linked environment might look, schematically, for Donne’s Holy Sonnet commonly known by the first half of its first line, ‘Oh my black soul’: Here we see two transcriptions at the left margin, one of a particular copy (British Library, shelfmark Vet.A2e363) of a document (or manifestation), which is the 1633 edition of Donne’s *Poems* named with the siglum ‘A’ by the Donne Variorum and numbered ‘78’ by Keynes, ‘S121864’ by the ESTC, and ‘STC7045’ by Pollard and Redgrave (STC). Here, we are interested in a work that is never named in existing library infrastructure, a



**Figure 7:** Sample entry from Robert's *Bibliography* (2004, 270).



**Figure 8:** A schematic diagram of linked resources for HSBlack.

poem known as *HSBlack* following the Donne Variorum conventions, which is expressed as ‘item 13’ in document ‘A’ (in Keynes78, to which I have added the item number 13 in the diagram). So, for the purposes of our data structure for a Donne archive, Keynes provides a crucial piece of information: an added, granular level of naming. On the other side of the schematic, linking to the EST entry number provides another layer of information, pointing to individual copies of the document that contains this item expressing the work ‘HSBlack’ (here I name only Vet.A2e363). We also have a transcription of a manuscript witness of this work (named item DnJ2845 in Beal) occurring in a document named ‘NY3’ by the Donne Variorum (commonly known as the Westmoreland MS). At the same time that this work, through a network of linked resources, links through to transcriptions, it also (moving in the opposite direction in the scheme) leverages John Robert’s naming conventions for linking to a secondary source (with an abstract provided by Roberts). Through Roberts, which aims to be exhaustive in identifying all secondary literature related to Donne, and which extensively indexes Donne’s works (as covered in this secondary literature), we have a wide open horizon of possibilities for linking primary and secondary sources.

While there are crucial gaps in naming supplied by some of these resources (most crucially, Beal), the conventions are in place and, when linked together, provide a sufficient level of granularity to enable a complex linked knowledge environment for the study of John Donne. Just how this data model might look when implemented, or how far such a sociable data structure could be generalized beyond this example remains to be seen.

## Competing Interests

The author has no competing interests to declare.

## References

- Beal, Peter. 2013. *Catalogue of English Literary Manuscripts 1450–1700*. Accessed August 1, 2017. <http://www.celm-ms.org.uk/>.
- British Library. n.d. *The English Short Title Catalogue (ESTC)*. Accessed August 1, 2017. <http://estc.bl.uk>.
- Coyle, Karen. 2016. *FRBR, Before and After: A Look at Our Bibliographic Models*. Chicago: American Library Association.
- Donne, John. 2005. The Variorum Edition of the Poetry of John Donne. *The Holy Sonnets*, Gary A. Stringer (ed.), 7(part 1). Bloomington, IN: Indiana University Press.
- Donne, John. 2013. The Oxford Edition of the Sermons of John Donne. Peter McCullough (ed.). Oxford: Oxford University Press.
- Gadd, Ian. 2009. “The Use and Misuse of Early English Books Online *Literature*.” *Compass*, 6(3): 680–92. DOI: <https://doi.org/10.1111/j.1741-4113.2009.00632.x>
- IFLA Study Group on the Functional Requirements of Bibliographic Records. 1998. “Functional Requirements of Bibliographic Records: Final Report.” München: K. G. Saur. Accessed August 1, 2017. <http://www.ifla.org/VII/s13/frbr/frbr.pdf>.
- Keynes, Geoffrey. 1973. *A Bibliography of Dr. John Donne*. 4th ed. Oxford: Oxford University Press.
- Marotti, Arthur F. 1986. *John Donne, Coterie Poet*. Madison, WI: University of Wisconsin Press.
- McGann, Jerome J. 1992. *A Critique of Modern Textual Criticism*. 2nd ed. Charlottesville, VA: University of Virginia Press.
- McKenzie, Donald Francis. 1999. *Bibliography and the Sociology of Texts*. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511483226>
- Nelson, Brent. 2013. “Radiant Donne: A Case for the Digital Archive and the John Donne Society’s Digital Prose Project.” *John Donne Journal: Studies in the Age of Donne*, 32: 175–200.
- Nelson, Brent. 2014. “From Index to Interoperability: The Desideratum of Authority Files in Large-Scale Digital Projects.” *Scholarly and Research Communication*, 5(4): 1–13. DOI: <https://doi.org/10.22230/src.2014v5n4a192>
- OCLC. 2017. “OCLC Research Activities and IFLA’s Functional Requirements for Bibliographic Records.” *OCLC Research*. Accessed August 1, 2017. <http://www.oclc.org/research/activities/frbr.html>.
- Pollard, Alfred W., and Gilbert Richard Redgrave. 1976. *A Short-Title Catalogue of Books Printed in England, Scotland, & Ireland and of English Books Printed Abroad, 1475–1640*. 2nd ed. rev. London: Bibliographical Society.
- Roberts, John R. 1973. *John Donne: An Annotated Bibliography of Modern Criticism, 1912–1967*. Columbia, MO: University of Missouri Press.

- Roberts, John R. 1982. *John Donne: An Annotated Bibliography of Modern Criticism, 1968–1978*. Columbia, MO: University of Missouri Press.
- Roberts, John R. 2004. *John Donne: An Annotated Bibliography of Modern Criticism, 1979–1995*. Pittsburgh, PA: Duquesne University Press.
- Roberts, John R. 2013. *John Donne: An Annotated Bibliography of Modern Criticism, 1996–2008*. Greenville, NC: DigitalDonne. <http://donnevariorum.tamu.edu/html/resources/robertsbib/Roberts1996.pdf>.
- Robinson, Peter. 2015. "Chapter 7: Social Editions, Social Editing, Social Texts." *Digital Studies/Le champ numérique*, 6. DOI: <https://doi.org/10.16995/dscn.6>
- Robinson, Peter. 2016. "Project-Based Digital Humanities and Social, Digital, and Scholarly Editions." *Digital Scholarship in the Humanities*, 31(4): 875–889. DOI: <https://doi.org/10.1093/llc/fqw020>
- Siemens, Ray, et al. 2012. "Toward Modeling the Social Edition: An Approach to Understanding the Electronic Scholarly Edition in the Context of New and Emerging Social Media." *Literary and Linguistic Computing*, 27(4): 445–461. DOI: <https://doi.org/10.1093/llc/fqs013>
- Smiraglia, Richard P. 2001. *The Nature of "a Work": Implications for the Organization of Knowledge*. Lanham, MD: Scarecrow Press.
- Stahmer, Carl. 2016. "Making MARC Agnostic: Transforming the English Short Title Catalogue for the Linked Data Universe." In: *Linked Data for Cultural Heritage*, Ed Jones, and Michele Seikel (eds.), 23–40. Chicago: ALA Editions.
- Stringer, Gary A. 2011. "The Composition and Dissemination of Donne's Writings." In: *The Oxford Handbook of John Donne*, Jeanne Shami, et al. (eds.), 12–25. Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/oxfordhb/9780199218608.013.0003>
- Stringer, Gary A. (ed.) n.d. "DigitalDonne: The Online Variorum." Accessed August 1, 2017. <http://digitaldonne.tamu.edu/>.
- Welsh, Anne. 2017. "From WEMI to WI to WII: FRBR, BIBFRAME and the 21st Century Bibliographic Model." *Catalogue and Index*, 186: 20–29. Accessed June 7, 2018. [https://archive.cilip.org.uk/sites/default/files/media/document/2017/ci186welsh\\_anne.\\_from\\_wemi\\_to\\_wi\\_to\\_wii\\_frbr\\_bibframe\\_the\\_21st\\_cent.pdf/](https://archive.cilip.org.uk/sites/default/files/media/document/2017/ci186welsh_anne._from_wemi_to_wi_to_wii_frbr_bibframe_the_21st_cent.pdf/).
- Wing, Donald G., and Alfred W. Pollard. 1972. *Short-Title Catalogue of Books Printed in England, Scotland, Ireland, Wales, and British America, and of English Books Printed in Other Countries, 1641–1700*, 2nd ed. New York: Index Committee of the Modern Language Association of America.

**How to cite this article:** Nelson, Brent. 2019. The Sociable Textual Archive: Laying the Groundwork for Linked Bibliographic Entities. *KULA: knowledge creation, dissemination, and preservation studies* 3(1): 4. DOI: <https://doi.org/10.5334/kula.12>

**Submitted:** 05 August 2017

**Accepted:** 17 October 2018

**Published:** 27 February 2019

**Copyright:** © 2019 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

 *KULA: knowledge creation, dissemination, and preservation studies* is a peer-reviewed open access journal published by Ubiquity Press.

**OPEN ACCESS** 