Radiant Donne:
A Case for the Digital Archive and the
John Donne Society's Digital Prose Project

Brent Nelson

What would Donne have thought of the computer? In *The Renaissance Computer: Knowledge Technology in the First Age of Printing*, Neil Rhodes confidently declares that John Donne would have loved it![1] Whether a Donne of the twenty-first-century might have written to his friend Sir Henry Wotton, "Sir, more than kisses, Facebook mingles souls," we can't really say, but we can say that for scholars of Donne, computers now mingle with almost everything we do. If we were to pursue further a connection between the age of Donne and our own, it might be to suggest that we take Donne's example in facing boldly the new advances in knowledge and technology that are fundamentally changing the academic world as we know it. Donne recognized that his world was on the threshold of fundamental technological and epistemological change, and in the face of this change he was at once fascinated and skeptical, curious and cautious. It was not so much communication technology that interested Donne, but rather "new philosophy call[ing] all in doubt." For scholars of Donne, it is the advancing dawn of the digital age calling us to new ways of understanding and engaging our discipline in one of its most foundational aspects: textual scholarship. At the 24th annual meeting of the John Donne Society in Baton Rouge, 19–21 February 2009, the

---

[1]Rhodes, "Articulate Networks: The Self, the Book and the World," in *The Renaissance Computer: Knowledge Technology in the First Age of Printing*, ed. Neil Rhodes and Jonathan Sawday (London: Routledge, 2000), pp. 184–196.

executive committee of the society adopted the John Donne Digital Prose Project. This essay is presented in a Donnean spirit of boldness as a kind of aubade, but not to resist the dawn, nor to rail against it, but rather to welcome what could and should be a radiant Donne of the digital age, and to do so on our own terms.[2]

The titular metaphor of Jerome J. McGann's *Radiant Textuality: Literature After the World Wide Web* is suggestive of two distinct applications to textual scholarship in the digital age. First, in the vein of post-structuralist theory, McGann's solar metaphor alludes to the capacity of digitally remediated documents to expose further the polyvalence and proliferation of meaning inherent in imaginative literature: the more you look at it, and the more ways you look at it, the more a text radiates meaning. Digital tools enable new powers of observation and analysis, thus more light. The second aspect of textual radiance is the ability afforded by the electronic medium to bring a constellation of reading tools and supporting material into close orbit of the text. These include representations and simulations of the materiality of text, scholarly annotation and contextual information, related and associated documents, and even text manipulation tools. A key aspect of this latter expression of radiance derives from the malleability and, McGann insists, instability of the digital text: its ready opportunity for "deformance," that is, for re-formation and re-presentation in ways that make it seem strange, dislocating it from traditional and habitual readings and forcing upon the imagination new ways of seeing it.[3]

In many ways, Peter L. Shillingsburg's *From Gutenberg to Google*: *Electronic Representations of Literary Texts* exemplifies McGann's suggestive notion of digital textuality in envisioning what a scholarly edition or repository of primary materials might look like. The basis of Shillingsburg's imaginings is what he calls *script acts*, which (drawing on speech act theory) he defines as "every sort of act conducted in relation to written and printed texts, including every act of reproduction and every

---

[2]Although was written under the aegis of the executive committee of the John Donne Society, with valuable input from the project's advisory board, I assume final responsibility for the ideas expressed in this essay.

[3]McGann, *Radiant Textuality: Literature After the World Wide Web* (New York: Palgrave, 2001), pp. 101–102, 104ff. McGann develops this notion of "deformance" (performance+deformation) together with Lisa Samuels.

act of reading."[4] Similar to McGann in his focus on the instantiated text, Shillingsburg insists on representing a literary work in all of its iterations, productions, and reproductions as disseminated and received through time and changing contexts.[5] The World Wide Web presents new possibilities for facilitating this interplay of text and context.

McGann and Shillingsburg evince a great deal of optimism as they imagine the possibilities of editing in the electronic medium, and yet this call to revolution in editing can be the cause of some anxiety. As Julia Flanders writes, "The technical developments that have emerged in the past two decades have encouraged the wholesale relocation of scholarly editing onto the digital platforms."[6] This move, however, is not value-neutral. "[T]his shift of ground," continues Flanders, "has also heightened awareness in the scholarly community of the reciprocal resonance of medium and editorial epistemology."[7] In the digital realm, a text is no longer fixed and singular, but extensible, alterable, combinable, and endlessly improvable (and, it must be admitted, corruptible and vulnerable to annihilation), and almost limitless in comparison with the constraints dictated by print publication. This seemingly new notion of text in the digital medium is really an extension of larger trends and debates in scholarly editing, such as the movement to "unedit" the text and to attend instead to the various material instantiations of a work of literature in its native forms.[8] On the other end of the editing spectrum,

---

[4]Shillingsburg, *From Gutenberg to Google*: *Electronic Representations of Literary Texts* (Cambridge: Cambridge University Press, 2006), pp. 40, 45.

[5]Shillingsburg, pp. 54, 67.

[6]Flanders, "Introduction" to a special issue on "Computing the Edition," *Literary & Linguistic Computing* 24.1 (2009): 1–8. Flanders's introduction and the collection as a whole provide a good introduction to the issues and opportunities of editing in the electronic medium (although most of the essays in this special issue are based on papers first presented over ten years ago).

[7]Flanders, p. 2.

[8]For a good summary of historical trends in editing and the recent developments in the materiality of the text in the context of "un-editing," see Leah S. Marcus, *Unediting the Renaissance: Shakespeare, Marlowe, Milton* (London: Routledge, 1996), pp. 1–5, 17–31. Randall McLeod has observed the power of the photo-facsimile to liberate the text from accumulated projections and interventions of meaning ("UN Editing Shak-speare," *Sub-Stance* 33/34

Flanders sees the digital edition as the next logical step in the modern development of scholarly editorial practice:

> [T]he emergence of the modern form of the critical apparatus, as an effort to systematize a body of information on sources and variants, seems to aspire to the condition of data and in its fullest form anticipates the modern database in its rigorous structures and the analytical power [it] afford[s] the scholar. Seen as an extension of this history, modern digital editing takes advantage of technologies that emphasize analytical systems in order to realize these aspirations—putting the editor in possession of the strongest possible tools for making sense of the complex textual landscape.[9]

In the digital medium, these two seemingly contradictory notions of editing can sit comfortably together. If it had been initiated a decade later than it was, the *Variorum* itself, one of the early implementers of a digital collation tool, might have been able to take full advantage of the electronic medium to represent what is perhaps the apogee of the critical edition—in particular, to find unforeseen solutions to the challenging task of representing in a complete way the complex textual and critical history of Donne's poetry.[10] Complementing the print edition is the *Variorum*'s DigitalDonne site, which aims to present digital representations—both facsimiles and machine-readable transcriptions—of various historical instantiations of Donne's poems.

This essay presents the case for an extension of Donne's digital presence to include the full corpus of his prose works. But why do we

---

[1982]: 37). On the context of digital editing, see Raymond G. Siemens, "Unediting and Non-Editions," *Anglia* 119.3 (2001): 423–455.

[9]Flanders, p. 1.

[10]Gary Stringer, gen. ed. *The Variorum Edition of the Poetry of John Donne* (Bloomington and Indianapolis: Indiana University Press, 1995–). See Brian Vickers's review of volume 8 in *Analytical & Enumerative Bibliography* n. s. 10.2 (1999): 107–111 for an assessment of the *Variorum*'s editorial achievement. Daniel O'Donnell makes the case that the modern critical edition provides an efficient and effective model for structuring textual data and a starting point for developing new ways of editing complex texts in the digital medium ("Back to the Future: What Digital Editors Can Learn from Print Editorial Practice," *Literary and Linguistics Computing* 24.1 [April 2009]: 113–125).

need a full digital treatment of Donne? First, there is a decided cultural shift toward digital, open-access publication in academe.[11] The Atkins Report on "Revolutionizing Science and Engineering through Cyberinfrastructure" boldly announces that "a new age has dawned" in science-based research.[12] The same could be said of the humanities. In a report on "The University's Role in the Dissemination of Research and Scholarship—A Call to Action," four national American associations of higher learning, including the Association of American Universities (AAU), highlight the importance of this cultural shift in the new digital environment:

> Decades of investment and development in information technologies and networked information resources have created an unprecedented opportunity for scholars to express, document, organize, and transmit knowledge with extraordinary flexibility, depth, and power; these same developments have made it possible for this knowledge to be accessible throughout our society and globally at manageable costs.[13]

For the AAU and its partner organizations, this new opportunity is a call to embrace the possibilities of open access.[14] Free and public access in

---

[11]For a summary treatment of this cultural shift, see Nick Jankowski, "The Contours and Challenges of e-Research," in *e-Research: Transformation in Scholarly Practice*, ed. Nick Jankowski (London: Routledge, 2009), pp. 1–18.

[12]Daniel E. Atkins, Chair, "Revolutionizing Science and Engineering through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure," January 2003: 31 <http://www.communitytechnology.org/nsf_ci_report/report.pdf>.

[13]The Association of American Universities, Association of Research Libraries, The Coalition of Networked Information, and National Association of State Universities and Land Grant Colleges, "The University's Role in the Dissemination of Research and Scholarship—A Call to Action," February 2009: 3 <http://www.arl.org/bm~doc/disseminating-research-feb09.pdf>.

[14]"The University's Role in the Dissemination of Research and Scholarship—A Call to Action," p. 5. Other recent calls for open access include John Willinsky, *The Access Principle: The Case for Open Access to Research and Scholarship* (Cambridge, MA: MIT Press, 2006); and Gary Hall, *Digitize this Book!: The*

digital publication is increasingly expected by national academic funding agencies. The Social Science and Humanities Research Council of Canada (SSHRC) is actively promoting open access in the dissemination of the research it funds, the first major step being the Synergies Project, a national infrastructure for publication of open-access journals.[15] There is also a growing sense within the academy itself that our primary materials ought to be digital. McGann has argued that the entirety of our cultural heritage will need to be re-edited in order to take advantage of the new possibilities of the digital medium.[16]

Second, Donne is a particularly interesting candidate for digitization. Despite Donne's central status in the canon of early modern writing in English, he has never received a full and consistent treatment in a collected works, despite the long and important history of textual studies in Donne scholarship. This is not surprising, given not only the size but also the generic diversity of his corpus, from polemic to lyric poetry, from private devotions to Latin prose satire, covering equally diverse subjects, including canon law, theology, suicide, and geo-politics to name but a few. This variety is itself a compelling reason for a digital archive of reliable, scholarly texts that would provide unprecedented access through his entire corpus and present exciting opportunities for diachronic analysis of his entire literary output. Moreover, Donne's uncommonly full biography places him in an extensive network of social relations that touch on many of the cultural and political movements and trends of his time. In short, Donne's life and letters are uncommonly full and varied, making him a very good candidate for the expansive, multi-faceted treatment that a digital environment affords.[17] Interest in editing and re-

---

Politics of New Media, or Why We Need Open Access Now (Minneapolis: University of Minnesota Press, 2008), which focuses on the humanities.

[15]For a statement on SSHRC's policy on open access, see <http://www.sshrc-crsh.gc.ca/about-au_sujet/policies-politiques/open_access-libre_acces/index-eng.aspx>; on the Synergies Project, see <http://www.synergiescanada.org/page/about>.

[16]McGann, "Literary Scholarship in the Digital Future," *The Chronicle Review* [*The Chronicle of Higher Education Section 2*] 13 December 2002: B7. See McGann also on the need for literary scholars to take charge of digitizing of their primary materials.

[17]Kenneth M. Price provides a good summary of the range of possibilities in what can be included in a digital scholarly edition ("Electronic Scholarly

editing Donne remains strong, with the ongoing *Variorum* project and new editions of the sermons and prose letters in the works. In Donne studies there is also a very strong tradition of close reading and textual analysis. Both of these areas of scholarship can be well applied in digital media.

Third, we need a fully digitized corpus of Donne's writing because our libraries are already being digitized, and so far the agenda of digitizing has been largely determined by commercial interests without adequate consideration of the needs of the most exacting and invested readers of literature: the professional academics. As Ray Siemens and others in the Implementing New Knowledge Environments (INKE) have argued, humanities scholars—the most demanding users of text— need to be setting the agenda for the development of digital tools for reading and research, both within the academy and in the larger public sphere. If the new day of the ebook is upon us, and if Donne is to be a part of this new era of radiance, scholars of Donne had better make sure this radiance has a solid center: the text. The dynamic nature of the digital medium need not diminish in any way the foundational importance of the scholarly text. Both McGann and Shillingsburg emphasize the open-endedness and extensibility of the digital text, thus the malleability and transformability of the text in the electronic form. Every time a new edition is produced in print, it needs to be recomposed; and then it is fixed, so that the next edition must begin the whole process afresh. A digital edition never has to be finished, never outdated. It can always continue to radiate. But the center, the basic materials—i. e., the texts—must be solid. The rest of this essay surveys what can be done with a digital text, and then assesses the Donne materials currently available in the emerging universal digital library, before concluding with a plan for electronic publication of Donne's prose and, ultimately, a complete DigitalDonne.

<p style="text-align:center">*   *   *   *</p>

Electronic texts serve two main functions: access and analysis. Most users value electronic texts for their accessibility, and there are some

---

Editions," in *A Companion to Digital Literary Studies*, ed. Ray Siemens and Susan Schreibman [Malden, MA: Blackwell, 2007], pp. 434–450).

functions of access that electronic texts serve very well. First, there is access to the texts themselves in cases where they are remote and therefore logistically difficult to access. Freely accessible facsimile images and transcriptions of rare books and manuscripts have afforded equal opportunity to those who are not close to archives where such materials are held. Second, the digital medium enables access to elements within a text or a corpus of texts, first in locating, then in processing and visualizing textual elements of interest. The electronic text has an advantage if, for example, I want to locate a passage which I remember only imprecisely or for which I don't have an author or title, or if I am looking for a word or phrase in an author or text and have no particular passage in mind. Having located my text, if I want to quote a long passage in an essay, or include a short text in a course reading package, the cut-and-paste capabilities afforded by electronic texts are a welcome time saver. We should note, however, that electronic texts are not equal in every way to the printed book with respect to ease of access. If I know exactly the passage I am looking for, the print volume on my shelf presents some advantages in accessibility. When I want simply to read or to spend time thinking through a passage, especially if I happen to be on a beach or in the midst of a power outage, electronic books don't perform as well as do printed books. There have been advances in making electronic reading devices more portable and ergonomic, but they still have a long way to go before they will be as pleasurable or comfortable as a good old paper artifact. Still, for most scholars, the simple ability to do ctrl-f word searches is a great help in the routines of scholarship.

But there is more to be done. The second main function of the electronic text is for text analysis. Here there are two main sub-types. The first is close-reading analysis, and the second is corpus analysis. In the field of English Renaissance literature, there are a few terrific examples of texts that enable the former, but only a relative few, and these are still in progress. A digital edition of George Hebert's *The Temple*, edited by Robert Whalen and Christopher Hodgkins, is newly published by University of Virginia Press.[18] The electronic version of the *New Variorum Shakespeare* (eNSV) under the editorship of Paul Werstine

---

[18]Whalen and Hodgkins, eds., *The Digital Temple: A Documentary Edition of George Herbert's English Verse* (Charlottesville: University of Virginia Press, 2012).

for the MLA is using an interface designed by Alan Galey to enable the user to manipulate the way in which its complex textual apparatus is displayed to the user: this work of translating the print text into electronic form is just beginning.[19] The Internet Shakespeare Editions, edited by Michael Best at the University of Victoria, is providing a complete set of digital-born scholarly editions of Shakespeare's writing, but again, this project in still in progress.[20] In a similar vein, the *Variorum*'s DigitalDonne site offers transcriptions (with press variants) and facing page facsimiles of the first three editions; the St. Paul's, Westmoreland, and O'Flahertie manuscripts; and the 1654 prose letters, together with concordances for the poems and letters.[21] But, such examples of texts that are reliable and resourceful enough to recommend themselves for close-reading are few. That is, we need accurately transcribed and edited texts that meet scholarly standards, *and* we need added features and functions that take advantage of the special capacities of the digital medium (these are what digital humanists like to call "affordances").[22] Encoding a text in XML using the TEI (the Text Encoding Initiative) guidelines, for example, enables the text to be transformed into a variety of formats and to be indexed in complex ways to enable, for example, retrieval of every person, text, or place mentioned in a work. This would, in effect, enable new ways of looking at the text.

The second sub-function is corpus analysis, that is, analysis across a large body of text or texts that would otherwise be difficult or impossible using the naked eye in standard methods of reading. Concording is the most familiar tool of this kind, but digital concordances can go beyond simply compiling lists of word references, a function that print

[19]Werstine, "Past is Prologue: Electronic New Variorum Shakespeares," *Shakespeare* 4.3 (2008): 208–220.

[20]See <http://internetshakespeare.uvic.ca/index.html>. Other initiatives include Digital Renaissance Editions: Early Modern Drama Online, ed. Brett D. Hirsch <http://www.digital-renaissance.info/>.

[21]See <http://digitaldonne.tamu.edu/>.

[22]Susan Hockey, *Electronic Texts in the Humanities: Principles and Practice* (Oxford: Oxford University Press, 2000), pp. 3, 132ff. For an evaluation of the degree to which current electronic editions have achieved these added affordances, see Lina Karlsson and Linda Malm, "Revolution or Remediation? A Study of Electronic Scholarly Editions on the Web," *HumanIT* 7.1 (2004): 1–46 <http://etjanst.hb.se/bhs/ith//1-7/lklm.pdf>.

concordances can do equally well. They might also provide instances of selected words within a definable context; combinations of words within a defined proximity (collocates); or words with a common root, prefix, or suffix. Other tools provide statistical analysis: the average number of words per paragraph, or letters per word, or the occurrences and distribution of rare words.[23] It is not difficult to imagine how such analysis might be applied across Donne's corpus. In his 1992 essay on Donne's consultation of the church fathers, Mark Vessey bases his argument on a sampling of references: "It would be instructive," he suggests, "to extend this kind of analysis of Donne's patristic references to a larger sample of his sermons."[24] A complete body of sermons (including marginalia), with even shallow encoding, would easily return every passage in which one of the church fathers is mentioned within a definable context. Text analysis tools would not only enable their location but also report the distribution of these references throughout the sermons. One can easily imagine the benefits of such a resource in extending and building on Katrin Ettenhuber's recently published study of Donne's use of Augustine to examine Donne's full range of patristic sources.[25] Analysis across the corpus of Donne's sermons could also track changes in style and attitude through Donne's preaching career. It could also help answer some bibliographic questions. In his essay on "The publication of John Donne's sermons," Robert Krueger concludes with the hypothesis that the *Fifty Sermons*, published by John Donne Jr. in

---

[23]Martin Mueller provides an excellent introduction to the uses of corpus analysis for literary study in "Digital Shakespeare, or towards a Literary Informatics," *Shakespeare* 4.3 (2008): 284–301, especially pp. 291ff. For a more general introduction to the methods and techniques of electronic text analysis, see Svenja Adolphs, *Introducing Electronic Text Analysis: A Practical Guide for Language and Literary Studies* (New York: Routledge, 2006), pp. 2–8, 37–50 (chapters 3–4 provide elaboration of some of these techniques). For a sample of text analysis tools, see TAPoR (the Text Analysis Portal for Research) at <http://portal.tapor.ca/portal/portal> and the most recent iteration, Voyant <http://voyant-tools.org/>.

[24]Vessey, "Consulting the Fathers: Invention and Meditation in Donne's Sermon on Psalm 51:7 ('Purge me with hyssop')," *John Donne Journal* 11.1-2 (1992): 106.

[25]Ettenhuber, *Donne's Augustine: Renaissance Cultures of Interpretation* (Oxford: Oxford University Press, 2011).

1649, were transcribed by two of the senior Donne's servants, Robert Christmas and Thomas Roper, and that the *LXXX Sermons* of 1640 were prepared "with great care" for the press by either Donne himself or Henry King, to whom Donne entrusted his papers.[26] Statistical analysis of the original printings of Donne's sermons, either comparing these two folios along the lines Kreuger suggests, or perhaps comparing sermons printed immediately or soon after they were preached with those that were published years later, could tell us a great deal about the composition and transmission of Donne's sermons. With even deeper tagging of morphological and lexical units, a scholar could conduct complex grammatical and linguistic searches to identify trends in an author's style.[27]

<div align="center">*   *   *   *</div>

There is much that could be accomplished digitally with Donne, but at present we don't have the most basic materials to which to apply these tools: that is, we don't have the texts. In some respects, this is a result of the awkward and tentative cultural shift in our discipline toward a digitized research environment. A divide remains between the standard scholarly editing projects that aim, in the first instance, to produce reliable, scholarly, critical editions of the highest standard for close textual analysis by the naked eye, and "digital" texts whose first priority is computer-assisted search and analysis across large corpora. These two functions are rarely integrated. Inspired by the promise of the Carnegie Melon Million Books Project, Gregory Crane, in his introduction to a 2006 special issue of *D-Lib Magazine*, asked the question, "what do you do with a million books?"[28] At twenty million books scanned as of March

---

[26]Krueger, "The Publication of John Donne's Sermons," *RES* n. s. 15 (1964): 159–160. Kreuger is refuting Potter and Simpson's assertion that the two volumes of sermons "should be regarded as essentially one volume" and that they derived from manuscript copies of "the same excellence" (citing George R. Potter and Evelyn M. Simpson, eds., *The Sermons of John Donne*, 10 vols. [Berkeley: University of California Press, 1953–1962]: 1:5, 46–48, 80–81).

[27]See WordHoard <http://wordhoard.northwestern.edu/userman/index.html>.

[28]Crane, "What Do You Do with a Million Books?," *D-Lib Magazine* 12.3 (March 2006) <http://www.dlib.org/dlib/march06/crane/03crane.html>.

2012, Google Books is the most notorious and prolific of the mass-digitization projects, which also include the likes of the Open Content Alliance (OCA), the World Public Library, the Internet Archive, and Project Gutenberg, among others.[29] But now that we have millions of digitized books, what can we do with them? While digital utopians such as *Wired* magazine's Kevin Kelly champion this new universal library and the "single liquid fabric of interconnected words and ideas" that it promises, many are concerned about how this library is being produced and what sort of library is being delivered. The gravest concerns are over Google's effective monopoly in building this "universal" digital library, evidently with the intention of commercializing it.[30] Moves have been made to break this monopoly in the name of open-access; nonetheless, Google, a profit-driven corporation, is largely responsible for this proliferation of digitized books online in such an aggressive manner that they are effectively establishing *de facto* what the public understand a digital book and the digital universal library to be.[31] As scholars, we might also ask what sort of "book" comprises this library. The enticement of what can be done with a huge corpus of books has resulted in a push toward universal coverage at the cost of completeness and correctness:

---

[29]Jennifer Howard, "Google Begins to Scale Back Its Scanning of Books From University Libraries," *The Chronicle of Higher Education* 14 May 2012 <http://chronicle.com/article/Google-Begins-to-Scale-Back/131109/>.

[30]Kelly, "Scan This Book!" *New York Times* 14 May 2006 <http://www.nytimes.com/2006/05/14/magazine/14publishing.html?_r=1&emceta1&pagewanted=all>. Kelly provides an excellent description of the intricacies, including copyright and economics, involved in producing this library. See also Cory Doctorow, "Google Book Search Settlement Gives Google a Virtual Monopoly over Literature," *boingboing* 17 April 2009 <http://boingboing.net/2009/04/17/google-book-search-s-1.html>; and Timothy B. Lee, "Federal Judge Rejects Google Book Monopoly," *ars technica* 22 March 2011 <http://arstechnica.com/tech-policy/2011/03/judge-rejects-google-book-monopoly/>. For a general critique of large digitization projects, see Kalev Leetaru, "Mass Book Digitization: The Deeper Story of Google Books and the Open Content Alliance," *First Monday* 6 October 2008 <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/2101>. See also Robert Darnton, "Google & the Future of Books," *The New York Times Review of Books* 12 February 2009 <http://www.nybooks.com/articles/22281>.

[31]Maggie Shiels, "Tech Giants Unite against Google," *BBC News Online* 21 August 2009 <http://news.bbc.co.uk/2/hi/technology/8200624.stm>.

there may be a million books, but from a scholarly point of view, a great many of these are "bad" books.

To highlight the need for a complete corpus of reliable texts, I would like to examine what is currently available to a scholar of Donne's prose, beginning with what is perhaps the most familiar and valued electronic resource, the Brigham Young University (BYU) electronic facsimiles of the Potter and Simpson edition of the sermons. The first and most obvious limitation is that the site provides none of the supporting textual apparatus of the original, print publication. Most importantly, the page facsimiles it provides are stripped of the marginalia, which appeared in the first printings of these sermons, as well as the line numbering supplied by the modern editors. These sermons, therefore, are neither complete nor correct. They do not serve well as reading texts. Neither do they function as a convenient cut-and-paste shortcut for incorporating quotations into articles or lecture notes, even though the page facsimiles are matched with OCR-produced machine readable text. The opening text of *Deaths Duell*, for example, reads thus:

> 2 PSA 68 VERS 20 IN PINE FINE AND ND UNTO WTO amm nto GOD THE LORD BELONG THE ISSUES OF DEATH ie FROM DEATH **DUILDINGS** standbythebenefit oftheirfoundations that susteine and support them and of their butteresses butteresses that comprehend y and embrace them and of their contignations that knit and unite them the foundations suffer them not to sinie sinke the butteresses butteresses suffer them not to swerve and the contignation and knitting suffers them not to cleave the body of our building is in the former part of this thi thls verse it is this hee that is our god is the god of salvation ad salutes of salvations salvations in the plurall plurals plurall so it is in the originall originale originall the god that gives us spirituall spirituals and temporall temporale temporall salvation too but of this **build- ing** the foundation the butteresses butteresses the contignations are in this part of the verse which constitutes our text and in the three divers accey **accep lations tations** of the words amongst our expositors expository expositors vnto anto god the lord be- long the issues of death for first the foundation of this building that our god is the god of all salvations salvations is laid in this that unto this god the lord belong the issues of death that is it is in his power to give us an issue and deliverance even then when wee are brought to the jawes and teeth of death and to the lippes of that whirlepoole whirlepoole the

> grave and so in this acceptation this exitus mortis this issue of death is liberatto liberatio liberatio a morte a deliverance from death and this is the most obvious and most ordinary acceptation of these words and that upon which our translation **lales** hold the issues from death. . . .

What this reveals is that the OCR scan has not been proofed: it is dirty OCR, as all mass-digitized OCR is.[32] The text needs a good deal of brushing-up to be useful for even the most basic reading purposes, and thoroughly proofed before it is fit to be re-used in a document of any kind.

The purpose of this digital resource, then, must be accessibility. The indexes serve this function, but only imperfectly. Most obviously, the index and the display of results do not include volume and sermon numbers to correlate with Potter and Simpson, but more fundamentally, the index itself is faulty. For example, in the "audience" index, "royalty" provides fifteen results, but omits one sermon for which the title explicitly states that the king was in attendance, the famous *Deaths Duell*, and misses another, Donne's first extant sermon, preached, as Peter E. McCullough argues, at the Palace at Greenwich where the king was *probably* present.[33] The second omission is understandable, given that the index was probably generated automatically by searching the title headings. The omission of *Deaths Duell* is more surprising, given that the

---

[32]On the "noise" (metaphorically, the unwanted elements that pollute a clean text) of OCR in mass digitization, see Crane, "What Do You Do with a Million Books?" For an early attempt at assessing OCR accuracy in large-scale projects such as Google Books, see Shaolei Feng and R. Manmatha, "Hierarchical, HMM-based Automatic Evaluation of OCR Accuracy for a Digital Library of Books," in *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries, Chapel Hill, NC, 2006* (New York: ACM, 2006), pp. 109–118 <http://portal.acm.org/ citation.cfm?id=1141753.1141776>. Standard practices in OCR are still unable to cope with the vagaries of early modern printing, although significant improvements are on the horizon. The Early Modern OCR Project at Texas A&M (eMOP), with major funding from the Andrew W. Mellon Foundation, is working to improve OCR and supporting techniques for producing machine readable text from early modern print sources (see <http://emop.tamu.edu/>).

[33]McCullough, "Donne as Preacher at Court: Precarious Inthronization," in *John Donne's Professional Lives*, ed. David Colclough (Cambridge: D. S. Brewer, 2003), pp. 179–204.

sermon's header in Potter and Simpson explicitly states that the king *was* in attendance. The index also omits every sermon preached at Whitehall whose header does not explicitly name the King as being in attendance (despite the fact that the index specifies, at the top-left corner of the screen, the Boolean terms that were used to produce the delivered results: King OR Queen OR Prince OR Whitehall).[34] It makes sense to include this group of sermons, except in cases where external evidence confirms that the King for certain was not in attendance. Most of these Whitehall sermons were Donne's traditional Lenten sermon, which the king customarily attended. The index also includes one sermon preached at Whitehall "to the Household," a specification that seems to exclude the king.[35] A user who comes to this index expecting to find sermons preached to the nobility will miss at least two sermons for which one of the nobility was present and several more for which they were almost certainly present. Erroneous inclusions (such as the sermon to the household at Whitehall) can be filtered out by the sophisticated user, but in a search tool, omissions escape only the most skeptical and scrupulous users.

But what of the meat and potatoes of text-based searches, the keyword search? Here again, the user will get incomplete results. Consider again the opening passage of *Deaths Duell*. Someone searching for instances of the word "building," for example, will locate a couple of occurrences, but miss two. This is a problem with the basic quality of the underlying text (in the sample above, uncorrected errors in OCR are underlined and bolded). Moreover, if you want to search for buildings, you had better make sure you enter every possible variation of the root word (including plurals); or if you are searching for "buttress," you need to anticipate the irregular spelling used here: "butteress." A user should eventually catch-on to the limitations of the search mechanism and devise complex Boolean searches to compensate for the lack of standardization in orthography, but no user can anticipate truncated results owing to dirty OCR. Although it is unlikely in this case, because "building" occurs correctly elsewhere in the passage, it is disconcerting to

---

[34]In the current interface, the terms are simply listed; in the previous interface, the Boolean logic was made explicit.

[35]Potter and Simpson make the same assertion, 7:8.

think that a search for architectural imagery might not turn up *Deaths Duell* among its results.

What else does the "universal digital library" have to offer a Donne scholar? Anthony Grafton reports a conversation with a Cambridge University librarian who told him that "[c]onservatively . . . 95 percent of all scholarly inquiries start at Google."[36] This being the case, we should like to know what Google Books has to offer a scholar interested in architectural imagery in Donne's sermons.[37] A search for "Deaths Duell" OR "Death's Duel" returns only one result: an 1840 edition that parcels the sermon with *Devotions Upon Emergent Occasions*. A search for only the regularized title—"Death's Duel"—provides a few more citations. One of them, a repackaging of the 1959 Ann Arbor edition of *Devotions* with *Death's Duel*, provides only page facsimiles of the sermon. You can't do much with this document except keyword searches and page-by-page viewing (Google books does not provide full two-page openings). You can't select and copy text; you can't even print a page. The book is non-portable, frozen intractably on the computer screen. (In this case, the printed book is much more versatile!) The other available editions provide only part of the text. All of them use modernized spelling.

The most generous Google "book" to provide a text of Donne's sermons is a digitized version of Henry Alford's edition of 1839. This is the type of book that forms the bulk of the Google Books corpus: a copy-clear nineteenth-century edition, an edition that no scholar would use as a principal text for research.[38] Alford's edition, for instance, is

---

[36]Grafton, *Worlds Made by Words: Scholarship and Community in the Modern West* (Cambridge, MA: Harvard University Press, 2009), p. 299.

[37]For a case-study examination of Google Books from a scholar's perspective, see Paul Duguid, "Inheritance and Loss? A Brief Survey of Google Books," *First Monday* 6 August 2007 <http://firstmonday.org/issues/issue12_8/duguid/index.html>.

[38]Dino Buzzetti and Jerome McGann illustrate this problem with reference to a widely used database in English studies—Chadwyck-Healey's English Poetry Full-Text Database (600–1900)—and similarly conclude that "from a scholar's point of view, this work is primarily an electronic concordance for the authors and works in question. While the texts have been for the most part carefully proofed, they are nearly all noncopyrighted" ("Critical Editing in a Digital Horizon," in *Electronic Textual Editing*, ed. Lou Burnard, Katherine

modernized in spelling and punctuation and slightly emended and
bowdlerized.[39] In this case, the user is presented with the text in three
forms: a page-by-page reading interface, a downloadable pdf of the entire
volume in facsimile, and a plain text version rendered through OCR.
The plain text is provided in batches amounting to four pages each in
Alford's original. Although Google's OCR is more accurate than that of
the BYU sermons, there are problems with the text: "issms" for "issues";
"woiiib" for "womb"; "de,ath" for "death," all within the first couple of
pages. Finally, if the scholar is interested in other sermons containing
building imagery, he or she might be out of luck: only four of Alford's six
volumes are currently available on Google Books.

Probably the best source for digital texts of Donne's writing is Early
English Books On-line (EEBO), which offers through its Text Creation
Partnership (TCP) almost all the works in Donne's corpus. The most
obvious limitation of this source is its restricted access: only a
comparative few libraries can afford to subscribe to EEBO, and even
fewer to the TCP and the transcriptions it offers. Another major
limitation is, again, the text. Although TCP claims "99.995% character
accuracy," the texts are presented explicitly as "reading" texts and are not
intended to meet a scholarly standard. In the case of *Deaths Duell*, there
are a handful of substantive errors and many more incidental errors in
capitalization, italicization, and punctuation.[40] There are many instances
of lacunae where the transcribers could not decipher the text of their
exemplar, including two full, consecutive pages. If we were to interpret
our example topic generously to include what might not be considered
architecture *per se*, a scholar seeking building imagery would again be
cheated of a passage with potential relevance, where Donne refers to

---

O'Brien O'Keefe, and John Unsworth [New York: Modern Language
Association of America, 2006], p. 59).

   [39]For a thorough examination of the historical importance of this edition,
despite its inadequacies in scholarship, see Dayton Haskin, *John Donne in the
Nineteenth Century* (Oxford: Oxford University Press, 2007), pp. 33–36, 58–67.
Haskin clarifies that only one sermon was bowdlerized, and only in two places
(p. 61).

   [40]Substantive errors include "Iesus" for "Jesus," "yeas" for "years," "fariendi"
for "faciendi," "contract act" for "contract," "consider ation" for "consideration,"
"excessueius" for "excess eius," and—in the appended elegy—"griewd" for
"greiu'd" and "camist" for "cam'st."

death as "the *gate* of *my prison*" which might "be *opened* with an *oyled key*," "*hewen downe*" or "*burnt downe*."[41] What we do have with EEBO's TCP are texts that can be used by any member of its subscribing institution as a basis for a new edition, and as the bases for the transcriptions in our own archive, but they are not themselves of archive quality.[42]

So, what can we do with this sampling of the million books available on the Internet? For scholars, not enough. In the case of some Google "books," we can't even cut and paste, print, or "capture" the text in any way. At least a printed book can be photocopied. In the case of EEBO we can search across a large corpus of Donne's works, and thousands more, and these texts are easily transportable for manipulation and analysis using digital tools, but the results of this research will be suspect owing to incompleteness and unreliable quality. EEBO is particularly powerful if we are doing contextual searches (e. g., for architectural references), but not for careful textual and statistical analysis of Donne's corpus.

This brief survey of Donne's corner of the universal digital library makes clear the need for something better. This library is too important to leave in the hands of non-expert readers, or automated processes and machines. In the world of mass digitization of books, when "standards" are evoked, they typically pertain to file formats and computer processes, not to the resulting text itself.[43] In scholarship, the question of standards applies in the first place to the quality of the texts we use. In what has become a foundational document in digital humanities, John Unsworth provides a list of what he calls "scholarly primitives."[44] It is not intended

---

[41]Potter and Simpson, 10:241.

[42]Under "Licensing and Access" in "Project Description / Goals & Strategies," Early English Books Online Text Creation Project <http://www.lib.umich.edu/tcp/eebo/proj_des/pd_intro.html>. In supplying these transcriptions, the TCP is invaluable to projects such as ours, and Paul Schaffner at TCP and Aaron McCollogh before him, have been eminently responsive to our requests for transcriptions and permission to use them.

[43]Karen Coyle, "Mass Digitization of Books," *The Journal of Academic Librarianship* 32.6 (November 2006): 641–645.

[44]Unsworth, "Scholarly Primitives: What Methods Do Humanities Researchers Have in Common, and How Might Our Tools Reflect This?," presented at a symposium on "Humanities Computing: Formal Methods,

to be an exhaustive list of basic scholarly activities, but the list as is can be taken as a fundamental set of functions that any electronic text should anticipate in its end product: discovering, annotating, comparing, referring, sampling, illustrating, representing. But there is perhaps even a prior primitive to these. Before any of these functions can be applied to a literary text, the scholar must be confident in the quality of that text. This is the first point in the Modern Language Association's "Guidelines for Editors of Scholarly Editions": "The scholarly edition's basic task is to present a reliable text: scholarly editions make clear what they promise and keep their promises. Reliability is established by accuracy, adequacy, appropriateness, consistency, explicitness."[45] With respect to reliability, there should be no distinction between our printed and electronic texts.

<center>*   *   *   *</center>

Given this critique of what is currently available, the burden now falls upon the John Donne Society's Digital Prose Project (JDSDPP) to offer something concrete that answers these deficiencies. Here is an opening gambit in this direction. The JDSDPP is really a first-phase measure to develop the basic digital materials for Donne's prose, and these will in future be integrated with the *Variorum*'s materials to form a complete DigitalDonne archive. The Digital Prose site will be everything one expects of a digital archive. It will include transcriptions of every significant witness of each of Donne's prose works, digital facsimiles of the source documents, digital surrogates of landmark materials related to Donne, integrated tools for textual analysis, bibliographies of primary and secondary materials, and provision for user contribution. And, of course, all this material will be linked. This will be a massive venture with no clear end in sight: ideally, it will never be finished, but will continue to grow with the addition of new materials and resources as they become available.

Experimental Practice" at King's College, London, 13 May 2000 <http://www3.isrl.illinois.edu/~unsworth//Kings.5-00/primitives.html>.

[45]MLA Committee on Scholarly Editions, "Guidelines for Editors of Scholarly Editions," The Modern Languages Association <http://www.mla.org/cse_guidelines>.

The intention here is not to give a complete plan for development of a resource that will be years in the making, during which time new possibilities and practices in digital approaches will develop and replace current standard practices. The plan will, by the very nature of its chosen medium, always be evolving. What I want to focus on here are some of the basic principles that will determine the kind of resource this will be and how it will be executed.

1. It will be open-access, freely and readily accessible to all who have an Internet connection. There might be spin-off projects or aspects of the archive that fall under copyright control, but the basic materials will be subject to a creative commons licensing agreement.[46]

2. As a scholarly project, it will be committed to the high standards of quality and accuracy that scholars have come to expect from their resources.

3. The Digital Prose Archive will be subject to rigorous peer review and editorial oversight. Best practices are being developed for assessment of digital resources but remain in flux. Our own practices will thus develop (and we plan to be part of the ongoing conversation), but to start, we will adopt the following mechanisms:

> a) An advisory board that will offer advice on matters of policy and practice and facilitate peer-review. This board is now in place.

> b) An editorial board that will manage the work of the project. At present, immediate oversight of the project is provided by me as director/general editor and Lara Crowley as associate director/editor. Each document will also be assigned an editor who will provide instruction and oversight of every participant's work in the preparation of that document.

---

[46]See Creative Commons <http://creativecommons.org/licenses/>.

c) Individual objects—such as transcriptions, bibliographic records, metadata, and images—will require specialized assessment, and we will rely on our advisory board for direction on how each assessment should be conducted and for help in implementing it. At each milestone of the project (e. g., the completion of a complete set of prose transcriptions), the entire archive will be subject to arms-length review. We plan to participate fully in the Renaissance node (REKn) proposed for the Advanced Research Consortium (ARC, which now comprises NINES [Networked Infrastructure for Nineteenth-Century Electronic Scholarship], 18thConnect [Eighteenth-century Scholarship Online], and soon MESA [Medieval Electronic Scholarly Alliance]) and will use the peer-review services it offers.[47]

4. Every participant's contribution and the nature and extent of that contribution will be duly attributed. This attribution will be part of the meta-data attached to each object.

5. This archive will be rich in meta-data, meaning that each digital object (transcription, image, or article abstract, for examples) will have attached to it a full representation of bibliographic, archival, and editorial information. This requirement is best satisfied with current technology using RDF (Resource Description Framework) headers with URIs (Uniform Resource Identifiers) assigned to the materials.[48] This metadata will ensure that the object can be found by anyone searching for it on the Web and that related objects can be linked to it.

---

[47]ARC is an extension of the NINES model (see <http://www.nines.org/>) to include all digital resources in the humanities covering the middle ages up to the modern era (see <http://liberalarts.tamu.edu/html/news-texas-a-m-s-college-of-liberal-arts-to-house-digital-literary-research-consortiu.html>).

[48]"RDF is a standard model for data interchange on the Web" (Working Group, "Resources Description Framework (RDF)," W3C <http://www.w3.org/RDF/>).

6. We will seek every opportunity to plug into available scholarly infrastructures to make the resource of the archive visible and easily accessible to those who might have an interest in or need for its materials. These elements of infrastructure will include such institutions and services as the university library, WorldCat, Google Books, the Library of Congress and other national repositories, and scholarly portals such as Iter.[49] This is the first cornerstone of the archive's plan for preservation and sustainability. Copies of the raw materials will be deposited in partnering institutional repositories and digital libraries, and the user interfaces (there could potentially be more than one) will be hosted by partnering institutions.

7. The archive will be built by a community of interested users. This is not the project of a closed society. To be sure, the viability of this project depends on the commitment of the John Donne Society, with its long tradition of collaboration and involvement of its membership, but it will also be a public project that will welcome and indeed seek the participation of any interest party as far as he or she is able. It will, in other words, welcome the new possibilities of crowd-sourcing in recognition that there are many potential participants outside the society, and even outside academia, who have a serious interest in Donne and his writing. In keeping with point two above, all contributors and contributions—professional and citizen scholar alike—will be subject to strict editorial oversight (see further below in relation to "Textual Communities"). Community involvement is the second cornerstone of the archive's plan for sustainability. It will continue to be built and updated as long as the community of users stays interested and active in its development. While we recognize that some scholars might choose to continue developing materials independently for commercial publication, we will encourage everyone who develops primary and secondary resources to deposit their basic materials in the archive. This provision of open, collaborative involvement also pertains to the acquisition of materials, most notably high-quality digital images of original documents. The project is built on the assumption that there will be no major funding for the purchase of digital images and the rights to publish them from major research libraries. Instead, interested scholars will be

---

[49]See <http://www.itergateway.org/>.

asked to seek out materials in their own or partner institutions, or request purchase and digitization of new acquisitions, and, crucially, the images and rights to publish them.

8. The project must be able to go forward without major funding, though funding opportunities will be pursued wherever they may be found. The first point is crucial: many ambitious projects have fizzled or floundered because major funding could not be had or ran out. As digital projects proliferate, as they have already begun to do, funding will be increasingly stretched thin. Our approach for the project in general will be to ask stake-holders to seek support, either through institutional grants or support in kind, to supply resources, particularly human resources, for preparing the materials. To date, participants at Texas Tech University, the University of Northern Illinois, Seattle University, and the University of Saskatchewan have received tangible support for their work on the project, mostly for hiring student research assistants.

9. In the near term, while our first set of materials are in beta-form, hosting of the bulk of the project will continue to be provided by Texas A&M libraries and the Humanities and Fine Arts Digital Research Centre at the University of Saskatchewan. We have begun exploring options for depositing finished objects in an established archive of early modern resources.

10. The editorial team will work with individual scholars to support their work with the archive's materials and, in turn, will explore ways in which individual scholarly projects might contribute to the archive. Although our mandate at present is limited to the production of basic primary and secondary materials, under the guidance of the advisory board and approval of the executive committee of the John Donne Society, we will remain open to new possibilities for publication and resource development within the digital archive.

   Phase One of the John Donne Society Digital Prose Project. As indicated above, the first phase toward a full Digital Prose Archive will focus on the prose works of Donne to bring these materials to the same state of preparation as those of the *Donne Variorum*'s DigitalDonne. It will comprise a complete set of transcriptions of every witness of every prose

work by Donne. There are three aspects to this work: fresh digitization of exemplars (original documents); completion and correction of EEBO-TCP transcription against these exemplars; XML markup of the transcriptions using a simplification of the "TEI-light" guidelines, to make the markup process as simple and accessible to volunteer scholars as possible.[50] EEBO transcriptions are being corrected by volunteer participants against fresh digital images of the copy text. Many of the errors and omissions in the EEBO transcriptions are the result of poor quality images of the exemplar. Therefore, a key element of the JDSDPP is an ambitious digitization project to locate, photograph, and secure permission to publish fresh images of at least one copy of every major witness of Donne's printed prose and multiple copies as feasible. At present these are provided through a combination of partnerships and ad hoc arrangements with libraries at Texas A&M University, the University of Saskatchewan, St. Andrews University, the University of Alberta, the University of Victoria, and the University of Connecticut, as well as the Huntington Library, and by means of the generous digitization policy of the Fisher Rare Book Library at the University of Toronto. In many cases, we have been able to acquire these images through the advocacy of a colleague in his or her home institution (see point seven above).

We are now approximately halfway through checking and correcting transcriptions of the sermons. We are beginning with the sermons for three reasons: they are the most studied of Donne's prose (perhaps with the exception of the *Devotions*); as a large body of texts of the same genre spanning some fourteen years of his professional writing career, they are particularly amenable to the particular affordances of electronic text analysis; and they can easily be parceled out in discrete, manageable portions to multiple participants. At the same time, a few scholars have assumed oversight of the transcriptions of the shorter prose works. These include *Essays in Divinity* (Sean Davidson), *Ignatius his Conclave* (Sean McDowell), *Conclave Ignati* (Piers Brown), *Biathanatos* (Siobhan Collins), and *Paradoxes and Problems* (Lara Crowley). So far over two dozen volunteers, from graduate students to senior scholars, have participated in transcription checking and acquisition of digital images.

---

[50]"TEI Lite," Text Encoding Initiative <http://www.tei-c.org/Guidelines/Customization/Lite/>.

Building on this foundation of broad participation, our next step is implementation of a browser-based environment for managing the work of transcription, recording of metadata, and other activities related to digital editing. This suite of tools is being developed by Peter Robinson and the Textual Communities Project at the University of Saskatchewan supported by a grant from the Canadian Foundation for Innovation, and in collaboration with the Workspace for Collaborative Editing project at Birmingham, England, and Münster and Trier, Germany.[51] It will provide a framework for inviting contributors, professional and amateur, to participate as they are able and to be trained and managed to ensure the production of high quality materials. This framework will enable users to develop or incorporate other browser-based tools to add other functions to the workflow of digital editing, including a collation tool (under development) and mechanisms for generating and linking related materials.

Next Steps. Future phases are yet to be determined and will be subject to available resources (chiefly, participant enthusiasm for involvement) and subject to guidance of the advisory board and approval of the executive committee of the John Donne Society. For phase one, once a critical mass of peer-reviewed resources is in place, we will design a dedicated, integrated website for the archive. In the medium term, we will maintain a public blog for beta versions of our digital objects as a way of quickly disseminating materials as they become available. We will begin with *XXVI Sermons*, including a reader-friendly Web-version linked to page facsimiles, a corrected basic ASCII version, and the raw XML file. Soon to follow will be *Fifty Sermons*. This work can be accomplished with our current resources. Our next major body of work will be *LXXX Sermons*. Our current set of volunteer participants have carried the project a good distance, but we will need fresh hands and eyes to carry forward the work of transcription, transcription checking, XML

---

[51]The "textual communities" involved in the project in the first instance include the Canterbury Tales Project (Peter Robinson), the Magic Manuscripts project (Frank Klaassen), and the John Donne Society's Digital Prose Project (Brent Nelson). See the Textual Communities workspace <http://www.textualcommunities.usask.ca>. On the Workspace for Collaborative Editing project, see <http://www.vmr.birmingham.ac.uk/itsee/2011/07/09/workspace-update/>.

markup, and acquisition and digitization of primary documents. We welcome all interested parties to visit the project blog to explore opportunities for involvement.[52]

University of Saskatchewan

---

[52]See <http://www.donneprose.org>.