Digital Day Thoughts and the Interpretive Edition as Dataset

The inspiration for my digital project, *Scholars' Grotto* (<www.scholarsgrotto.com>), was a small, practical need. No text-based electronic version of *The Relief; or, Day Thoughts* (1754),[1] a parody of graveyard poetry by Henry Jones, appeared to exist, and I wanted to perform some basic lexical analysis. The poem is short enough for such work "by hand," but a digital transcription would be helpful if I ever needed to search the text later. The thought of producing a clean copy along the lines of Jack Lynch's eighteenth-century HTML texts sparked what we might call a digital-humanities chain reaction.[2] What if other readers could benefit from access? Would providing the text alone suffice? What about producing something citable? Could an encoding in TEI solve problems of quality, archiving, scholarly reliability? And what if I pushed beyond careful transcription to include features that promote critical analysis?

Such cascading questions ultimately provoked the conviction that the digital humanities need a model for the individual interpretive edition as dataset. The interpretive element may seem reasonable. All remediation is interpretation, though what I mean to emphasize is the conscious communication of analytic content through encoding and presentation.[3] The language of datasets, however, may seem foreign because datasets evoke big data—large corpora of text, restricted by genre, time period, format, and/or other scholarly interest, which are then subjected to various techniques of mining and statistical analysis.

Framing the individual edition as "dataset" does not objectify the content as an anonymous member of an aggregate, but personalizes it for close examination. Each dataset, or unique encoding and rendering of a text, presents a different experience, and each can be woven

---

[1] Hereafter referred to as *Day Thoughts*.
[2] <http://andromeda.rutgers.edu/~jlynch/Texts/>.
[3] For transcription as interpretation, see Peter Robinson's "Towards a Theory of Digital Editions" on the "questions of intention, agency, authority, and meaning" involved (114).

into traditional scholarly discourse as a thread of evidence to be evaluated and revised. Thus, instead of relying on a single scholarly edition or diplomatic transcription, close readers can test a plurality of models of the text, much as "macroanalysts" can tweak data-mining algorithms and archival criteria to produce different models of literary history. Thinking in terms of both "dataset" and "interpretive edition" frees scholars to vivify texts through various digital techniques of updating, manipulating, forking, remixing, re-contextualizing, and re-visualizing. The paradigm is at once forward-looking and grounded: it leverages the advantages of computing technology but also demands a firm grasp of classic—and in literary criticism, too often neglected—matters of bibliography and editing in service of literary analysis.[4] The question is not (or not only) "how good are your data?", but "how compelling are your data?"

Some scholars have called for this kind of work. Making "A Case for Heavy Editing," Amanda Gailey shows how tagging metaphors, textual and pictorial idioms, and even suggestive omissions, can capture content and context otherwise lost in bare transcription and broad archival searching (133-34, 136-38).[5] Julia Flanders goes further in her keynote address to the 2012 *Text Encoding Initiative* conference, envisioning a future for TEI as a tool not solely of "reproduction or remediation," but also of hybrid forms between edition and monograph.[6] The problem is resources: time to produce quality work, but especially the funding and technical infrastructure to

---

[4] It responds partially to Jerome McGann's desire that "textual and editorial scholarship, often marginalized in humane studies as a narrowly technical domain, should be shifted back to the center of humanist attention" (2).

[5] Gailey concedes that this encoding practice could be controversial, but argues that "contestable tagging would not severely limit the usability of the document and might seem a more viable possibility for projects directed by literary scholars" (132). On a practical level, it should be a minor matter to produce an XSLT stylesheet that would generate a copy of the XML with "contestable tagging" removed.

[6] Published online as "TEI and Scholarship (in the C{r|l}o{w|u}d)." See also Stephen Ramsay's conception of an algorithmic criticism which can "assist the critic in the unfolding of interpretive possibilities."

sustain it. Efforts like the *TAPAS* project, which aims "to provide TEI publishing and repository services at low cost to those who lack institutional resources,"[7] will be a boon to the field. But as *Scholars' Grotto* shows, it is already possible to publish and archive interpretive digital work in a meaningful way, on virtually no budget.

Day Thoughts and Digital Technology

In the midst of learning TEI markup by encoding Jones's poem, I discovered a human- and machine-readable version in the collections of the *Text Creation Partnership* (*TCP*).[8] The rationale for transcribing the poem having evaporated in a moment, I needed a reason to continue encoding, and this proved a saving grace: the project evolved from transcription alone to include bibliographic and semantic scholarship, giving the work new meaning and motivation. Ensuring clean data was the gateway. I noticed that the *TCP* edition had several lacunae, indicated by UTF-8 symbols such as lozenges (U+25CA), which affect the dependability of the text. This seemed strange because the *Eighteenth Century Collections Online* (*ECCO*) PDF I was consulting—and from which the *TCP* edition was supposedly derived—is perfectly legible. Then I noticed that the *TCP* version employed a completely different capitalization strategy. Finally, I saw it: their transcription is based on an octavo edition, while I was encoding the quarto.

To someone not formally trained in bibliography, the capitalization change was curious. English authors and publishers were, in the mid eighteenth century, starting to set substantives in lower case, but it was surprising to see the transition enacted in two editions of the same poem from the same year. Which should have authority? In an analysis of Alexander Pope's editing practice, David Foxon underscores the difficulty of determining who controls meaning in

---

[7] <http://tapasproject.org/about>.
[8] <http://name.umdl.umich.edu/004808620.0001.000>.

eighteenth-century publications by quoting John Smith's *Printer's Grammar* (1755, printed one

year after *Day Thoughts*): compositors are expected to ask "the Author, or Master," whether "the

old way, with Capitals to Substantives" ought to be practiced (qtd. 122). Short of possessing a

manuscript or proof (we have neither for *Day Thoughts*), it would be, as Foxon says of

punctuation, an "almost impossible task of trying to establish which changes are due to the

author and which to the printer" (121). And it is questionable whether Jones could have exerted

the editorial influence of a major poet like Pope.[9] Though he was successful for a time, Henry

Jones was an interloper in print society, one of the natural geniuses (a bricklayer by trade) who

caught the eye of eighteenth-century sponsors. Given this uncertainty about which printing best

expresses Jones's intentions, analysis of his poem ought to take into account both. The quarto

and octavo are unique experiences of *Day Thoughts*. If something as noticeable as capitalization

differs, what subtler changes might be lurking in each version, and might they shed light on

Jones's manipulation of graveyardist poetic tropes?

Thus commenced my experiment encoding in parallel segmentation, or recording

multiple textual witnesses in the same XML file without privileging one over the other.[10] Since

the majority of the poem's nouns now needed to be marked, I opted to scrap my nearly complete

transcription of the quarto, but the project was advancing. While I was not part of an institutional

project, I was learning the technology. I had the benefit of taking Texas A&M's "Programming

for Humanists" online course (http://www.programming4humanists.org), which helped me over

---

[9] Foxon writes, "it is becoming clear that many authors, like Pope, revised the accidentals of their work in proof. Different authors (and the same author at different times) will do so to a widely varying extent" (121).

[10] For an exemplary exploration of parallel segmentation, see Tanya Clement's article about her edition of Baroness Elsa von Freytag-Loringhoven's poetry in "Knowledge Representation and Digital Scholarly Editions in Theory and Practice."

the hurdle of transforming XML into web editions through XSLT.[11] And I had the proof of concept for work that could advance scholarship, a kind of edition that was digitally dependent and expressed an argument in the encoding as well as the presentation.

Parallel segmentation is a valuable enterprise for its own sake. It enables all kinds of comparative work that would be unwieldy, if not impossible, in print versions. If one wishes to focus solely on comparing editions, there are good digital platforms available. *The Versioning Machine*, for example, enables encoders "to compare diplomatic versions of witnesses side by side, allowing for images of the witness to be viewed alongside the diplomatic edition, and providing users with an enhanced typology of notes."[12] This functionality is extended in the *Juxta* project, which provides additional comparison tools, direct website integration, and a "Commons" web server where scholars can host their comparative editions.[13] In fact, thanks to the standardization of TEI, one can plug the XML of *Digital Day Thoughts* into either software package with little effort.

What I wished to highlight does not figure in *The Versioning Machine*'s or *Juxta*'s presentational schemes, however. I wanted to see differences between the quarto and octavo classed according to self-defined categories. Thus, in my XML of *Digital Day Thoughts* I assigned words and phrases special "analysis attributes," which allowed me to highlight changes in inflection, spelling, and punctuation. While I also tagged capitalization as a change, I did not include an analysis attribute with the value "letter-case," partly because this change is ubiquitous. My encoding therefore implicitly takes the stance that capitalization is the baseline

---

[11] Special thanks to Laura Mandell, Laura Estill, Matthew Christy, Violeta Ilik, Luis Meneses, Quinn Dombrowski, and David Rettenmaier for teaching the course and incorporating live input from off-campus participants.

[12] <http://v-machine.org/documentation.php>.

[13] <http://www.juxtasoftware.org/> and <http://juxtacommons.org/>.

difference between the texts, while the other changes sprinkled throughout are special cases, ripe

for isolation and analysis. Neither the baseline assumption nor the assignment of specific

attribute values is semantically neutral. Yet, this interpretive content would be lost on *The*

Versioning Machine and *Juxta*.[14] This is not to criticize either platform, which would have

worked well with a different kind of project, but to draw attention to the fact that software—even

TEI itself—affects meaning. "Plugging in" to someone else's code requires scrutiny.[15]

For *Digital Day Thoughts*, a modest, custom-built web application would offer control

without overwhelming my technical resources. It was not something to take for granted,

however: design decisions in HTML can be easily overlooked, especially when there is no

budget for professional artistic input. And as McGann warns, "most—nearly all—websites

created in HTML will not outlive their creators, and the duration of the materials may well be

much shorter even than that" (29). It is tempting to quit the field and argue that the XML is all

that matters, that XML is what will survive when web hosts go down, projects are defunded, and

new HTML and scripting standards come into existence. Indeed, I am arguing for critique at the

XML level. The web interface, however, is the user's first encounter with the text. For the vast

majority, it is the only encounter. Though *Day Thoughts* will likely not become a digital "best

seller," period researchers, Gothic enthusiasts, and that supremely important readership,

undergraduate students, may have recourse to Henry Jones. Offering an interface that richly

expresses the encoding, encourages intuitive interaction, and displays the edition with some level

of aesthetic consideration is essential (even if fragile from an archival perspective). And thanks

to efforts such as *Reclaim Hosting* (the home of *Scholars' Grotto*), it is possible to host a web

---

[14] Through "a terrible irony," Julia Flanders writes, the small TEI project in its current form is "unlikely to be published with tools and interfaces that will make the most of its rich markup."
[15] Clement's work is a good example of plugging into the *Versioning Machine* reflectively.

site with advanced programming control for very little money, and without direct institutional support.[16]

In the web version of *Digital Day Thoughts*, users encounter the poem front-and-center with minimal distraction. A box to the right displays metadata and technical detail—original edition information, electronic version number, links to various downloads of the quarto and octavo as well as the master XML—and it scrolls unobtrusively away as the user continues to read. The real meat of the project, however, is the "Text Toolbox" on the left-hand side, which follows the user down the poem. This toolbox allows one to change the base text dynamically from quarto to octavo and then highlight differences between them. When a certain kind of difference is selected, distinctions between the printings are marked in crimson while the changed text is extracted for further analysis. For example, if "Quarto" is selected as the base text, clicking "All changes" will redden the quarto in place and will also pop the octavo's differences into the right-hand margin so that instant comparison can be made. Switching "All changes" to "Orthographic changes" will restrict the differences to spelling, while flipping the base text to "Octavo" will reset the poem at its current position, allowing the user to push quarto differences into the right-hand margin. Thus, neither the octavo nor the quarto is accorded pride of place, as though one is the progenitor and the other a derivation.

What makes the text an interpretive edition, conducive for use as a close-reading dataset, is not parallel segmentation in itself, but the encoding of textual features on the backend (in the XML) combined with the ability to call them out on the frontend (in the web version). Scholars could follow the same design process to manipulate a single witness, highlighting, for example, certain kinds of conjugation, words with a shared etymological background, metrical feet—

---

[16] *Reclaim Hosting* (http://reclaimhosting.com/) was specifically designed for academics by Jim Groom and Tim Owens of the University of Mary Washington.

whatever he or she has encoded.[17] Suppose, for example, a literary critic wishes to make an argument based on patterns of verb usage in a poem. The critic-cum-editor, having tagged linking, transitive, and intransitive verbs, generates an electronic edition that she uses to develop an article for an academic journal. Instead of simply asserting that the poet tends to include a preponderance of linking verbs when trying to achieve a certain affect/effect, the critic links back to the dataset that helped her reach this conclusion. With a few clicks, the end user can instantly analyze the poem's verb patterns, including examples the critic has neglected to quote in her article.

The next interested scholar examines this edition and finds that transitive verbs seem to complicate the argument. This hunch is confirmed when a third scholar forks the XML in order to mark ditransitive verbs, thence finding significance in the poet's inclusion or omission of prepositions with indirect objects. A fourth scholar rejects the scheme of highlighting (di)transitive, intransitive, and linking verbs to emphasize subordination instead, rewriting the HTML transformation code to underline all sentences in which transitive verbs appear in relative clauses following a linking verb (as in, "he is a scholar who creates digital editions"), making yet another turn to the debate.

Artificial though the example may be, it demonstrates how interpretative-edition datasets could function as a hybrid version of "traditional" and digital scholarship. The scholarly debate itself, rooted in the semantics of a certain poet, remains comprehensible in journal articles, even if the XMLs disappear. The datasets, however, facilitate the debate in multiple ways. The

---

[17] The *Digital Archives and Pacific Cultures* project is a move in this direction. In their edition of *The Injured Islanders* (Gerald Fitzgerald, 1779), the editors have encoded TEI referencing strings (<rs>) in a fashion similar to my analysis attributes. Based on these referencing strings, their HTML edition allows users to highlight analytical "cultural interactions," such as "Conflict," "Bloodshed," and "Trade" (see <http://pacific.pitt.edu/InjuredIsland.html>).

original edition opens itself to criticism first on the level of display—the second scholar notices something different using the exact same data and interface. It then opens itself to criticism on the encoding level when the third scholar notes an absence in the verb tags and modifies the XML. Criticism migrates to the programming level when the last scholar finds something new to visualize by altering the XML transformation code. No version of the text supersedes any other; each is useful as an edition which presents the text and as a dataset which makes an argument.[18]

A danger of this model is a chaotic proliferation of texts, but standards of scholarly practice and good software design can maintain order.[19] Each well-formed TEI XML will document both the source of the text, whether physical or electronic, and the history of emendation. The third scholar's forked copy will not, therefore, be suddenly orphaned or invalidated should the first scholar release an updated version of her XML; the encodings lead back, like a trail of breadcrumbs, to the original. Provenance and attribution are baked in.

Scholars' Grotto seeks to reduce the confusion of editions further by following a paradigm of software versioning for texts (starting with a "Beta 1" release, and incrementing to "1.0," "1.1," and so forth).[20] This practice exploits the editing flexibility of digital technology while also offering stability: the scholar-editor can fix mistakes as often as needed while assuring users that they will always have a clear citation (and a backup of older versions) should they wish to quote a text. Computer programmers will recognize this as a simplified form of a "Distributed Version Control System" (DVCS), which is already deployed in high-end digital

---

[18] As Ramsay argues, any scholarly reading of a text gives us "not the 'original' text, but a text transformed and transduced into an alternative vision," and "algorithmic transformation can provide the alternative visions."

[19] It might be objected that an overflow of XML from humanities scholars would be an enviable problem.

[20] In keeping with clear citation, I should note that the version of *Digital Day Thoughts* referenced throughout this article is the "Beta 1" release.

archives. Christian Wittern sketches a model of a rather advanced DVCS for scholarly editions that can accommodate widespread collaboration on translations, emendations, annotations, etc, using Git. He is guided by editorial interests, envisioning a workflow in which multiple scholars revise or extend the same work, either storing revisions on their own Git repository or committing them back to the main one pending authorization from the lead editors. Versioning need not be so complicated for interpretive editions, however, because datasets function as finite pieces of evidence for analysis. Where the Wittern model yields an ever-expanding, crowdsourced, but ultimately *singular* work, even when different branches emphasize different elements, the interpretive-edition model yields a discrete series of unique datasets, even when the scholars are forking the same source XML.

All versions of *Digital Day Thoughts* are maintained on the *Scholars' Grotto* web server and on *GitHub*, so that the entire revision history will live as long as either of the two web servers survive.[21] Most importantly, all versions can be downloaded to the user's computer, where they might live indefinitely. Web interface files are stored alongside the source XML, theoretically preserving the original experience for future emulation (or for direct access on a computer museum's ancient Windows 7 machine), while widely used, open web standards mitigate the danger of supersession in the first place. A visualization tool for Windows 3.11 or a web application optimized for Netscape Navigator might present problems today, but an older TEI SGML and its HTML 2.0 rendering could be manipulated, updated, and displayed with some effort and know-how. The rule of thumb on *Scholars' Grotto* is, if you can't download a

---

[21] See <https://github.com/scholarsgrotto/day-thoughts>.

digital artifact in human- and machine-readable form and manipulate it offline, it doesn't exist for practical scholarly purposes.[22]

This is not to discourage the development of closed systems and proprietary frameworks, per se, but to encourage a sustainable model for a certain kind of interpretive work. For the literary scholar working with individual texts, the preservation of the edition in as close to its original state as possible is of paramount importance. The first edition, the fourth and definitive edition, the 1912 anthologized edition, as well as 2014 digital editions are all important pieces of a text's history. Using freely available, broadly implemented, and carefully standardized software to create datasets that are downloadable from multiple sources and manipulable with simple text editors is the closest equivalent we have to archival-quality paper.[23]

Insights on *Day Thoughts*

Given the typesetting latitude afforded eighteenth-century compositors, the low social station of Henry Jones versus his high authorial pretensions,[24] and the absence of manuscripts, it is impossible to establish a best text for *The Relief; or, Day Thoughts*. Instead of ignoring the problem, *Digital Day Thoughts* encourages a proliferation of interpretations. Consider Jones's rebuke of Gothic credulity in the quarto printing: "Shall meer Negations, unsubstantial Shades, /

---

[22] Paige Morgan urges similar caution for the novice DH content creator: "If the platform doesn't allow you to download any file, or any file that can be read with an alternative program, then use the platform with extreme caution, for experimental purposes only" (<http://www.paigemorgan.net/how-to-get-a-digital-humanities-project-off-the-ground/>). Without citing individual examples, it is easy to imagine texts locked behind a commercial interface, comments lost on removed message-board software, markup features ignored because they require a login or a plugin download.

[23] See the "Acid-Free Bits" standards by Nick Montfort and Noah Wardrip-Fruin: <http://eliterature.org/pad/afb.html>.

[24] See William J. Christmas's account of Jones's attempts to fashion himself as a poet, and not a "bricklayer poet," in chapter 3 of *The Lab'ring Muses* (130-156).

Such Monsters form, to fright th' unthinking Crow / To Fancy tangible, to Terror real?" The imagery seems all of a piece, the macabre machinery of graveyard poetry proving fit carrion for the appetite of the degraded reader, who is aptly pictured as a "Crow"—a loud-mouthed scavenger, bedecked in black, associated with auguries and superstition. The crow is also a potential visual metaphor for vested clergymen: the immediate target of Jones's satire (singled out in the long title) is Church of England parson and graveyard poet Edward Young, author of *Night Thoughts*.

When comparing quarto to octavo, however, one notices that "Crow" becomes a "crowd," and after a moment's consideration, crowd does seem the likelier, if less exciting, candidate. A version of *Day Thoughts* in a 1756 collection of Jones's poetry, for example, adopts the octavo's reading of crowd, even though it retains the quarto's older style of capitalization. Yet one is loath to lose the colorful chain of associations conjured by the word "crow." The quarto is part of the poem's publishing history, widely available at a time when Jones was selling well. The quarto is the version glossed in the April 1754 issue of *The Monthly Review* (304), and "Crow" is the form preserved in the only modern reprinting of the poem, in volume two of the anthology, *Eighteenth-Century English Labouring-Class Poets 1700-1800* (Keegan 25, l. 51). It partakes in the sociology of the text and need not be emended away (nor offered as the definitive reading of the line).[25]

---

[25] See D.F. McKenzie's notion of bibliography as "the study of the sociology of texts" (5). McKenzie singles out misprinting as a particularly suggestive crux of textual sociology— William K. Wimsatt and Monroe C. Beardsley's misquotation of William Congreve's *The Way of the World* (1700) in "The Intentional Fallacy" (1946) provides a vivid example of how "misreading has become an historical document in its own right" (13; 11-19). As Robinson writes, "texts are embedded in complex webs of discourse, with multi-dimensional relations between author, text, everyone involved in the making and reception of a text, editor and audience" (107).

The octavo, meanwhile, reveals its own colorful readings. In the quarto, the narrator speaks of

> The Winds, that mournful yell, from ecchoing Vaults,
>
> And broken Sepulchres, their groaning Accents;
>
> As if they wail'd the long-departed Dead,
>
> Who slumber, deep in everlasting Night,
>
> Within these dreary Mansions;

The "inflectional changes" tool shows an octavo reading of "long-departing dead," however, which alters the valence of superstitious dread. In the quarto, the emphasis is on distance; what could be more ridiculous than mistaking the wind's "Accents" for a lament of the long-since deceased? In the octavo, however, the voice of the dead ambiguously mingles with the wind—as if the deceased are still in the act of departing and crying out in their "slumber." The target in both versions is superstitious folly, but instead of simply describing the mistake of the credulous, the octavo reproduces their frame of mind.

The 1756 edition opts again for the staid reading, "long-departed Dead," by changing allegiance to the quarto rendering. If *Digital Day Thoughts* were a rigorous critical edition, the proper method might have been to include this 1756 edition as another witness in the encoding— perhaps as *the* adjudicating witness. The point, however, has been precisely not to provide a "best text," but a highly malleable one, inclusive of suggestive changes.

One difference highlighted by the "Orthographic changes" button reveals a meaning shared by both versions that might have been lost on modern readers consulting either edition alone. After heaping up a collection of graveyard imagery for inspection, Jones in the quarto declares the whole a "detested Groupe! / A Landskip fit for Hell: the Work of Fiends!" Simple

enough, it would seem. When one notices that the octavo prints "detested growpe," however, the reader begins to smell significance. "Groupe" and "growpe," it turns out, are both older spellings of "groop," a livestock gutter. The twenty-first-century eye, glimpsing the old-fashioned form of "landskip" in the following line, is likely to scan over either "groupe" or "growpe" as an antiquated spelling of the more antiseptic "group."[26] Keegan's anthology, for example, neglects to gloss the unusual word (23, l. 41). Seeing the strange spellings next to each other, however, prompts a visit to the dictionary, which in turn raises further questions. Is Jones invoking antique orthography to tar superstitious graveyard poetry as a "Gothic" pursuit? Or could the two spellings be simple compositor choices? Might "groop," in either spelling, be a reflection of dialect from Jones's Irish or bricklaying past?[27]

The comparisons are suggestive on a number of levels, though it is fair to point out that weaknesses of my encoding strategy became apparent in the very process of encoding. Since capitalization changes are pervasive, the reader can become overwhelmed by the amount of information displayed with the "All changes" button, so that certain kinds of difference are

---

[26] Whether "landskip" would have been considered old-fashioned by eighteenth-century readers is its own interesting question. The *OED* lists the spelling as accepted usage in the fifteenth through seventeenth centuries, with a parenthetical determination of archaic for the eighteenth century, but according to the *OED*'s online interface, this entry has not been "fully updated" since first publication in 1901. Suggestively, "landskip" and "landscape" in Google's Ngram Viewer seem to duel for supremacy in the early eighteenth century, with "landskip" taking a temporary, spiking lead between 1751 and 1757, before "landscape" overwhelms thereafter. Even if the data were clean and dependable, however, the slight predominance of "landskip" in the 1750s would require its own extended analysis for an adequate explanation. Was it simply a widely-accepted variant? Were authors using the word ironically, in an antiquated sense? Were writers, for some reason, suddenly quoting the same older poet who had used the word "landskip"? See <https://books.google.com/ngrams/graph?content=landskip%2Clandscape &year_start=1600&year_end=1830&corpus=15&smoothing=3&share=&direct_url=t1%3B%2C landskip%3B%2Cc0%3B.t1%3B%2Clandscape%3B%2Cc0>.

[27] Since OCR of early-modern texts is undependable, and since "groupe" is also a possible variant spelling of "group" in the eighteenth century, the evidence of Google's Ngram Viewer is unhelpful for questions of orthography.

obscured. In one line, the quarto capitalizes the entire phrase "Common Sense," even though "Common" is an adjective, suggesting that Common Sense carries additional allegorical resonance which is lost in the octavo. It is reasonable to imagine Jones counterposing a god of reason to the devilry of superstition, and "Common Sense" might signal other allegorical resonances lost in the lower-casing of the octavo. But *Digital Day Thoughts* by default simply records the fact of capitalization change, one in a crowd of many.

To render this particular case of capitalization more visible, I assigned "common" and "sense" the attribute value "other changes," a kind of junk drawer approach to catching less categorically obvious differences. Are other semantically suggestive capitalizations hidden in the comparative text by the very proliferation of changes? The junk drawer will catch some but could obscure others, while another scheme of categorization might have illuminated the text in different ways altogether. This dynamic might seem disconcerting, but it is a particular strength of the interpretive-edition dataset; all readings of a work are partial, so there are always more avenues to be explored, which another dataset can render visible.

Digital Day Thoughts may not make as pointed an argument as the hypothetical transitive-verb datasets I posited earlier, but it does open several interpretive trails that radiate from bibliographic questions and terminate in the act of reading.[28] With tools that expose a wide array of changes between editions but also cull out a very specific selection of semantically rich changes, the dataset implicitly highlights a tension between vigorous, even unruly poetic idiom and regulatory editorial machinery. While all eighteenth-century authors had to negotiate the vagaries of publishing, the tension is especially telling in Jones's case. On the one hand, a more earthy idiom would be appropriate for a laboring-class poem—it could even be a selling point.

---

[28] "Every act of reading is in fact an act of critical editing," as G. Thomas Tanselle writes, though digital editions which allow dynamic manipulation of the text rather amplify the notion (6).

But Jones was endeavoring to excise the laboring-class sobriquet to become simply a poet. If there is a flattening of language, from "crow" to "crowd," or from long-departing to long-departed dead, Jones might have approved it as part of his professionalization. Ultimately, it is up to users to navigate the multitude of possible readings as they traverse *Digital Day Thoughts*, or to posit a new horizon of readings through their own markup of the text.

Conclusion—Toward a More Productively Boring Digital Humanities

By readily revealing the invested and contestable nature of its remediation, an interpretive-edition dataset calls for exploration of wider implications in another venue: the traditional modes of scholarly discourse. This may seem counterintuitive. Digital evidence naturally invites computational response, and it is important to develop infrastructure to support it.[29] But if the technical details of the dataset can be made to dissolve before its literary argument, even the most non-digital humanist might join the debate, especially if it is pursued in the pages of a favorite journal. "Analog" modes of intellectual exchange have developed over a very long time and continue to serve us with tested mechanisms of quality control, cachet, reliability, and sustainability. Perhaps introducing computational evidence into this system without a whiff of advertisement about its technological novelty will foster rapprochement with traditional humanities programs while showing that digital work is not just a set of inward-turning experiments conducted in a separate annex of the English department.

If the interpretive-edition dataset is particularly well suited for this rapprochement, it does not follow that it should supplant others kinds of digital editions, however. Editorially

---

[29] David Hill Radcliffe writes, "it seems conceivable that we blockheads in the humanities might, in addition to producing articles and monographs, begin constructing the digital infrastructure required to take eighteenth-century scholarship to the next level. Investing time and labor in this infrastructure is the right thing to do and will yield profits for years to come" (33).

focused texts, ranging from stripped-down digital facsimiles to full critical editions, will always be indispensable and distinct projects; the depth of the latter, in particular, far exceeds the scope of an interpretive dataset. In his recent article for *Textual Studies and the Enlarged Eighteenth Century*, James E. May explores the compelling stories that such detailed bibliography can reveal about textual creation and meaning while also demonstrating why, at its highest level, it is beyond the purview of unfunded literary scholars. May points to Jeffrey Reid Fox's 1980 dissertation edition of Young's *Night Thoughts*, "Night 1," as an example of a failure to find what turns out to be a multitude of variants despite Fox's having consulted six copies (65-66). Kit Kincade's inquiry into multiple university presses shows that Fox's six copies might have satisfied most publisher requirements, but it would fall short of more stringent expectations that editors consult up to twenty (123-127). From her own recent editorial work on the Stoke Newington Defoe, Kincade finds four copies of an edition a good starting base, "and if an anomaly is discovered, it has been my experience that an additional four are checked to help understand the nature of the anomaly" (125). This seems reasonable. But the specialization of bibliographic work, combined with the time and expense of visiting even a modest selection of between four and eight rare books rooms, points to the need for substantial financial backing of high-quality critical editions. The interpretive-edition dataset is simply a different species of digital edition, with different goals, different time investments, and a different funding model— which is not to say that solid editorial principles will be unnecessary; just that interpretation, not bibliography, is the main point.[30]

---

[30] For *Digital Day Thoughts*, I checked the *ECCO* quarto against a copy in the rare books collection of Princeton University. Curiously, the Princeton copy was bound with a miscellaneous collection of short poems and pamphlets. In a different kind of project, this rebinding of the poem would be another fascinating strand of bibliographic analysis to explore.

In the ideal digital world, critical editions and careful diplomatic transcriptions would serve as the source of interpretive datasets. The *Folger Digital Texts* project, which encodes down to the single-word level, would make an excellent "codebase" for forking interpretive editions of Shakespeare's plays, for example.[31] In fact, the MLA is experimenting with this kind of remediation in their second round of the "New Variorum Shakespeare Digital Challenge," offering scholars an XML of a play and inviting them to create "new means of displaying, representing, and exploring this data in the most exciting API, interface, visualization, or data-mining project."[32] This is a strong step toward the interpretive dataset model. The next step would be asking scholars to lard the XML and interface with interpretive content so that winners of the challenge could publish an accompanying article in *PMLA*.

Just as an individual, interpretive-edition dataset differs from a digital critical edition, so too a scholar's collection of datasets may look different from broader DH projects. Digital humanities archives tend to revolve around a thematic core—an author, a coterie, a time period, a genre, etc.—and this thematic core serves as a justification for the continuation of the project. With *The Women Writer's Project* or *The William Blake Archive*, the guiding principle is readily apparent.[33] With interpretive-edition datasets designed to be "plugged in," as it were, to traditional modes of scholarly discourse, the archive is bound to look different.[34] A scholar's web project, like the *Scholars' Grotto*, might host any number of eclectic datasets that offer fine readable editions, but the *raison d'être* of archive selection and text manipulation tools could

---

[31] <http://www.folgerdigitaltexts.org/>.

[32] <http://www.mla.org/nvs_challenge>, accessed August 1, 2014. One major caveat to the project should be noted: "TEI conformance is not possible because of the substantial number of structural adaptations necessary to accommodate NVS-specific features."

[33] <http://www.wwp.northeastern.edu/> and <http://www.blakearchive.org/blake/>.

[34] Although interpretive datasets can be incorporated into traditional thematic archives, as are the works in the *Digital Archives and Pacific Cultures* project (see note 17 above).

mystify until the user consults the creator's published scholarship.[35] This is not a bad thing. It may help render DH work uncontroversial, in a good sense—the disparate assemblage of datasets becoming no more noteworthy than would a listing of wide-ranging academic articles on a faculty member's webpage.

This productively boring diffusion of digital humanities work could lead to better integration of more "exciting" DH projects with traditional disciplinary research in the mid-term future, even if the number of active XML creators remains relatively low. Though the model of the interpretive-edition dataset reduces barriers to digital production by scaling back the requirements of a full critical edition and a well-funded archive, it doesn't remove them entirely; the creation of the barest TEI file and HTML transformation requires significant technological understanding. Some humanities scholars may feel they lack the prowess to learn XML, and a great many of those who could or do know it will be unwilling to spend the time, energy, and patience required to encode their argument into a dataset. It is fine if digital editions, large or small, remain niche work. But if the willing few can expand the accessibility of digital tools and conduct research in the channels of traditional humanities discourse, as "business as usual," it might go a long way toward demonstrating the value of the wider digital humanities ecosystem.

---

[35] Not that the published scholarship needs to be read—the datasets could spark insight completely unrelated to the creator's interpretive interests.

Works Cited

Christmas, William J. *The Lab'ring Muses: Work, Writing, and the Social Order in English Plebeian Poetry, 1730-1830*. Newark: U of Delaware P, 2001. Print.

Clement, Tanya. "Knowledge Representation and Digital Scholarly Editions in Theory and Practice." *Journal of the Text Encoding Initiative* 1 (2011): n. pag. Web. 15 Aug. 2014.

Fitzgerald, Gerald. *The Injured Islanders, or the Influence of Art upon the Happiness of Nature: A Poetical Epistle from Oberea of Otaheite to Captain Wallis*. Ed. Elisa E. Beshero-Bondar, et al. *Digital Archives and Pacific Cultures*. Web. 15 Aug. 2014. <http://pacific.pitt.edu/InjuredIsland.html>.

Flanders, Julia. "TEI and Scholarship (in the C{r|l}o{w|u}d)." *Bits that Matter*. 31 Aug. 2013. Web. 15 Aug. 2014.

Folger Digital Texts. Web. 15 Aug. 2014. <www.folgerdigitaltexts.org>.

Foxon, David. "Greg's Rationale and the Editing of Pope." *Library Journal* 5th ser. 33.2 (1978): 119-124. Web. 15 Aug. 2014.

Gailey, Amanda. "A Case for Heavy Editing: The Example of *Race and Children's Literature in the Gilded Age*." *The American Literature Scholar in the Digital Age*. Ann Arbor: U of Michigan P, 2011. 125-144. *digitalculturebooks*. Web. 15 Aug. 2014.

Jones, Henry. *The Relief; or, Day Thoughts: a Poem. Occasioned by the Complaint, or Night Thoughts. Humbly Inscribed to the Right Honourable the Earl of Holderness*. Octavo edition. London, 1754. *Eighteenth Century Collections Online*. Web. 15 Aug. 2014.

---. *The Relief; or, Day Thoughts: a Poem. Occasioned by the Complaint, or Night Thoughts. Humbly Inscribed to the Right Honourable the Earl of Holderness*. Octavo edition. London, 1754. *Text Creation Partnership*. Web. 15 Aug. 2014.

---. *The Relief; or, Day Thoughts: a Poem. Occasioned by the Complaint, or Night Thoughts.*

  Humbly Inscribed to the Right Honourable the Earl of Holderness. *Poems*. Dublin, 1756.

  Google Books. Web. 15 Aug. 2014.

---. *The Relief; or, Day Thoughts: a Poem. Occasioned by the Complaint, or Night Thoughts.*

  Humbly Inscribed to the Right Honourable the Earl of Holderness. Quarto edition.

  London, 1754. *Eighteenth Century Collections Online*. Web. 15 Aug. 2014.

Juxta. Web. 15 Aug. 2014. < http://www.juxtasoftware.org/>.

Keegan, Bridget, ed. *Eighteenth-Century English Labouring-Class Poets: 1700-1800*. Vol. 2.

  London: Pickering & Chatto, 2003. Print.

Kincade, Kit. "Editing Defoe: *A System of Magick* as a Case Study." *Topographies of the*

  Imagination: New Approaches to Daniel Defoe. Ed. Katherine Ellison, Kit Kincade, and

  Holly Faith Nelson. New York: AMS Press, 2014. 117-137. Print.

"Landscape." *The Oxford English Dictionary*. Online version of the 2nd ed. Web. 15 Aug. 2014.

  <http://www.oed.com/view/Entry/105515>.

Lynch, Jack, ed. *Electronic Texts*. Web. 15 Aug. 2014.

  <http://andromeda.rutgers.edu/~jlynch/Texts/>.

May, James E. "Threats to Bibliographic and Textual Studies Posed by Widely Distributed

  Filmed and Digitized Texts." *Textual Studies and the Enlarged Eighteenth Century*. Ed.

  Kevin L. Cope and Robert C. Leitz, III. Lewisburg: Bucknell UP, 2012. 61-97. Print.

McGann, Jerome. *A New Republic of Letters: Memory and Scholarship in the Age of Digital*

  Reproduction. Cambridge: Harvard UP, 2014. Print.

McKenzie, D.F. *Bibliography and the Sociology of Texts*. London: British Library, 1986. Print.

Montfort, Nick and Noah Wardrip-Fruin. "Acid-Free Bits: Recommendations for Long-Lasting

    Electronic Literature." *The Electronic Literature Organization*. June 2004. Web. 15 Aug.

    2014. <http://eliterature.org/pad/afb.html>.

The Monthly Review, or, Literary Journal. By Several Hands. Vol 10. London, 1754. *Google*

    Books. Web. 15 Aug. 2014.

Morgan, Paige. "How to Get a Digital Humanities Project off the Ground." Web. 15 Aug. 2014.

    <http://www.paigemorgan.net/how-to-get-a-digital-humanities-project-off-the-ground/>.

"The New Variorum Shakespeare Digital Challenge: The Second Round." *Modern Language*

    Association. Web. 15 Aug. 2014. <http://www.mla.org/nvs_challenge>.

Radcliffe, David Hill. "No Man but a Blockhead: What the Eighteenth Century Has to Teach Us

    about Digital Humanities." *Textual Studies and the Enlarged Eighteenth Century*. Ed.

    Kevin L. Cope and Robert C. Leitz, III. Lewisburg: Bucknell UP, 2012. 21-38. Print.

Ramsay, Stephen. "Algorithmic Criticism." *A Companion to Digital Literary Studies*. Ed. Susan

    Schreibman and Ray Siemens. Oxford: Blackwell, 2008. N. pag. Web. 15 Aug. 2014.

Robinson, Peter. "Towards a Theory of Digital Editions." *Variants* 10 (2013): 105-131. Web. 15

    Aug. 2014.

Tanselle, G. Thomas. "Editing without a Copy-Text." *Studies in Bibliography* 47 (1994): 1-22.

    Web. 15 Aug. 2014.

TAPAS Project. Web. 15 Aug. 2014. < http://tapasproject.org/about>.

The Versioning Machine. Ed. Susan Schreibman. Web. 15 Aug. 2014. <http://v-machine.org/>.

Wittern, Christian. "Beyond TEI: Returning the Text to the Reader." *Journal of the Text*

    Encoding Initiative 4 (2013): n. pag. Web. 15 Aug. 2014.