Abstract
This essay explores David Greetham’s notions of “textual forensics” in light of new forms of textual analytics practiced upon born-digital materials. It argues that computers and computational environments ask us to rethink basic evidentiary categories, i.e. “internal” vs. “external,” as well as such concepts as normality, agency, and intentionality in relation to textual criticism. In the process—through a forensic examination of one specific piece of digital media—we also learn something about David’s own personal computing habits.

for David Greetham

I didn’t attend the CUNY Graduate Center, so David Greetham was never my teacher in the classroom sense. Nonetheless, his textual voice has always been inseparable from that of my University of Virginia teacher, Jerome McGann. I don’t mean, of course, that they said the same things. Rather, that my encounter with these two giants of textual scholarship was collocated. In the fall of 1995 I was auditing McGann’s seminar, whose subject was nominally the Pre-Raphaelites but which really was the textual conditions of our ongoing encounter with the Pre-Raphaelite Circle. On the book list was David’s Textual Scholarship: An Introduction, just out the previous year from Garland, and still available only in cloth. The cost was something stratospheric to a graduate student, but I paid without hesitation, not only because Professor McGann said we had to, but because here, it was plain to me, was the essential counterpart to McGann’s polemic: the nuts and bolts of what we needed to know, at least to instruct us in how much we did not yet know. This book has remained close at hand on my shelf, and I have turned to it many times since.

While Jerry cut a local figure on the Grounds of the University of Virginia and the nearby Corner, D. C. Greetham seemed towering, remote. His own graduate classes, immortalized in the introduction to McGann’s Critique of Modern Textual Criticism with the tale of the improvised calypso sounded thrilling, but I couldn’t quite reconcile that carnival scene with
the intellectual presence that went on to publish, in very short order, monographs and collections such as The Margins of the Text, Textual Transgressions, and Theories of the Text, all within the next five years and all of which immediately found a place on my burgeoning bookshelves. At some point during this time, however, I attended my first STS conference, and the tall (he really was tall!) personage with the reading glasses round his neck who folded himself into one of the student desks in a mean little NYU classroom to hear my paper morphed from D. C. to just David. I did a global find and replace on my own internal hard drive, and it’s been David ever since.

Now I’ve admitted this in public before so it’s no great revelation here, but I sometimes think I’ve had exactly one really important idea in the course of my own scholarly work. This was the insight, such as it was, that the conversations then unfolding in the margins and pages of David’s work and elsewhere in the textual scholarship community were equally applicable to the conditions of electronic textual production. When I began working on that idea as a graduate student I sought confirmation wherever I could find it. It was there in D. F. McKenzie’s 1985 Panizzi lectures, where he explicitly included electronic data in his sociology of the bibliographical universe. It was certainly there in McGann, who always insisted that “hypertext” was the true subject of such books as he was then publishing, Black Riders and The Textual Condition. It was there in Random Cloud’s wandering writings, which overtly channeled information theory. And of course it was there in David’s work, both in concrete particulars such as the inclusion of an ASCII character table in the chapter on printed books in the Textual Scholarship volume, and in such contributions as “Is It Morphin’ Time”, a laser-sharp meditation on digital materiality by way of the Power Rangers which closed out a 1997 Oxford University Press collection on electronic textuality. But the work of David’s which most immediately served to ground my thinking, first in my dissertation and then in my first book, was an essay he contributed to a 1996 special issue of PMLA organized around the Status of Evidence. Now I know it may seem inconceivable to many of you that PMLA could actually devote one of its numbers to any subject quite so fascinating, but the truth is that this issue, edited by Heather Dubrow, is a marvel. It features a roundtable on the subject of evidence with W. J. T. Mitchell, Janice Radway, and David Vander Meulen, among others. The articles are diverse: T. Hugh Crawford offers a piece on medical imaging, and there is a detailed manuscript study of Charlotte Perkins Gilman’s “The Yellow Wallpaper”. David’s contribution was an essay entitled “Textual Forensics”.
The first thing one notices about David’s essay is it looks like a marvel. I mean that literally. It reproduces in facsimile the dramatic page scenes of Randall McLeod, whose own work is central to the discussion. There is a pleasing duality to the disruptive nature of this conspicuous display, for even as Randy is exploding the textualized norms of formal academic discourse, David, in reproducing these pages—are they image or text, illustration or (legible) textual appendage?—further complicates the staging of his own argumentation. “Textual Forensics” proceeds from the superimposition of the vocabulary between the forensic sciences and textual studies—notably evidence and witnesses—to take up questions related to the scientific method in bibliography and textual criticism, internal vs. external evidentiary states, and the conditions of bibliographical knowledge. But it is also a bold reconsideration and a repositioning of textual scholarship itself, which David dubs an antidiscipline, one that is both a postmodern pastiche of method and practice as well as lacking in any stable epistemological referent or even, he insists, an essential subject matter. Forensics itself, David reminds us, is a Janus-faced word, by definition both the presentation of scientific evidence and the construction of a rhetorical argument. Yet forensics for David arrives on the scene not primarily by way of
applied criminalistics, but rather through the notion of venatic lore, as presented by Carlo Ginzburg: marginal, seemingly insignificant details, mostly involuntary, which Giovanni Morelli used to authenticate the paintings of the European masters and which Ginzburg reads through both Freud and Arthur Conan Doyle as paradigmatic of a tradition of inference and deductive reasoning that David then brings to bear on the epistemology of textual knowing. This was a powerful lever for me as I began to ponder the accidental and substantives of computer-generated documents. What is the appropriate measure of intentionality, and what are its symptoms, in a textual environment where every coded signal—themselves always reducible to voltage differences, as a fundamentalist such as Friedrich Kittler would remind us—is the product of some procedural agency, be it human or algorithmically initiated? The vast majority of what is written on any computer hard drive is, after all, the product of the machine. If you doubt me on this I commend you to Diff in June, a recently available 1600-page volume documenting and recording every piece of data that changed on a single day in June on a single computer's hard drive. The project's initiator, Martin Howse, describes it as “a novel of data archaeology in progress tracking the overt and the covert, merging the legal and illegal, personal and administrative, source code and frozen systematics”. The pages, in other words, are a data dump, most of it simply opaque and even the infrequent pockets of legibility resisting any simple semantic engagement since they are rendered within the context of a now absent operating system.

My reading of “Textual Forensics” also dovetailed with my discovery of an applied field of practice known as computer forensics, defined by authorities as involving “the preservation, identification, extraction, documentation, and interpretation of computer data” (Kruse II and Heiser 2001, 1). Computer forensics has furnished the practical armature for what I believe are my most important engagements with both textual scholarship and digital humanities, and much of my work in Mechanisms consisted in aligning (with what success I leave it to you to determine) computer forensics as practiced by specialists with the precepts of textual scholarship as they were articulated by David, Jerry McGann, Don McKenzie, and others.

Computer forensics takes as its primary locus of investigation a specific class of digital object known as a disk image. “Image”, of course, is a commonplace term in computer network design, and refers to a perfect copy, or duplicate, of information at divergent points in the system. But image also carries with it the full freight of Western traditions of mimesis, from the inheritance of what W. J. T. Mitchell dubs iconology through the photographic revolution to the force of the facsimile image in modern editorial
practice, as demonstrated especially compellingly by McLeod. Likewise, as Heather MacNeil and Bonnie Mak remind us, the visibility of an image is deeply tied to notions of authenticity that derive from the function of records in evidentiary contexts: “The observational principles on which we ground our belief in records as trustworthy evidence [. . .] reflect a conception of records as witnesses to events, and a corresponding view of the world as one that is capable of being so witnessed” (2007, 40). Disk images obtained under appropriate conditions, including the use of cryptographic hashing, are legally acceptable as forensically sound substitutes for original storage media.

Consider this passage from the documentation of the AFF, or Advanced Forensic Format, detailing the function of one particular variable in the specification known as “badflag”:

The existence of the badflag makes it possible for forensic tools to distinguish between sectors that cannot be read and sectors that are filled with NULLs or another form of constant data, something that is not possible with traditional disk forensic tools. Tools that do not support the badflag will interpret each “bad” sector as a sector that begins with the words “BAD SECTOR” followed by random data; these sectors can thus be identified as being bad if they are encountered by a human examiner. Alternatively, AFF can be configured to return bad sectors as sectors filled with NULLs. (24)

In other words, at stake here is the investigator’s ability to discriminate among various levels of agency and intentionality in computational evidence. (In practice, one might be able to determine whether the contents of a particular file system have been deliberately tampered with.) But note too how the difference between evidence of absence and the absence of evidence is dependent on various acts of reading: what the disk imaging software can and cannot read from the physical media in question, and what the human investigator can or cannot read in the form of alphabetically encoded messages. The image is thus a site where signals are rendered as symbolic units (the “bits” of the bitstream, but also the hexadecimal and ASCII representations that one encounters with a typical viewer), all of which have varying degrees of semantic legibility. Put another way, the disk image is a site not just of mimetic imitation but also critical interpretation, based on the capabilities of both software and human analysts. Moreover, because a disk image is snapshot of the complete computing environment it effectively collapses the distinction between internal and
external evidence that David treats at length in his essay. Unlike a scholar such as McKenzie, who turns to the evidence of the archive to read the books—or a scholar such as Randy, who turns back to the books to ward away the pernicious influence of the overly edited archive—a disk image as evidentiary artifact is simply a linear stream of bits, text and context commingling in the one-dimensional topology of the string.

So let us now take a look at a textual body and the kind of evidence it reveals. Some weeks ago I contacted David and asked if he could find any of his own personal legacy digital storage media. He sent me two CD-ROMS, and as fortune would have it one of them contains a version of the “Textual Forensics” essay. Note that what we have here is not a forensically sound disk image in the manner I have just been describing, but rather a simple logical copy of a file system. Nonetheless, we’ll attempt a brief autopsy. First, we can get a sense of David’s directory structures and work habits as we navigate the CD. The file metadata, meanwhile, tells us David last touched this document very early in the morning of the last day of the year of 1997. Since the PMLA essay was in fact published in 1996 we can speculate as to his motives, but we can also consider that the essay may have simply been migrated to some new file system or media at that time.

Figure 2. “Publications / Toshiba 12 Aug 03”. Two CD-ROMS provided by Greetham. Photo by author.
Figure 3. Windows Properties of file “Textual Forensics Remnants” indicating such information as date last modified, date this instance of the file was created (presumably corresponding to the creation of the CD, above), file format, file size, and directory location.
In any case, opening it reveals a document entitled “Remnants of ‘Textual Forensics.’” The title notwithstanding, it seems something more like a draft, existing in some state prior to submission to the journal’s editors. Much of the text is consistent with what one reads in the published version, but there are variants throughout and it begins and ends in different places. Further inspection of the metadata tells me that this was originally a WPD or WordPerfect file, and that David was using a Lexmark Optra Plus laser printer at the time; given the mention of a Toshiba machine, almost certainly a laptop, on the CD itself, we can begin to reconstruct aspects of David’s computing environment that would be important to a future archivist seeking solutions to preserving this material. Here we see the rendition of the document in a hex viewer. The discontinuity between this and its presentation in my current copy of MS Word is jarring, but what is the definitive state of this digital artifact? It is no more this hexadecimal view than the seemingly normative presentation in my word processor, for both are in fact highly stylized renditions that are legible to us solely as the result of the imposition of various software logics, a phenomenon I

1. Greetham has subsequently suggested that this represents the first and fullest version of the essay, shortened at the request of PMLA, but its “remnants” retained in their original state for some possible future use (ultimately unrealized).
have elsewhere called formal materiality. Nonetheless, this particular view of David’s work, complete with the improperly rendered character codes (what we now call mojibake) has a patina of raw authenticity: these are the remnants of the .txtual body in perhaps their lowest state of legibility by any conventional means. To go further is to descend into the increasing abstractions of machine code and ultimately the pits and lands of the laser-scored surface of the CD itself.

“Where to stop? How to stop?” David asks rhetorically in his essay, echoing Foucault’s meditation on the extent of Nietzsche’s authorship, another key touchstone for him here (1996, 36). In the data dump that will characterize the textual scholarship of our very near future—recall the example of Diff in June, its printed bulk representing the capture of but a single day in the life of a system—this question will become all the more urgent. How much evidence is enough, and to what end when the archive itself consists of hundreds, thousands, even hundreds of thousands of variants, each date- and time-stamped to the millisecond? The textual forensics of the near-future will, I think, require its own forms of Big Data operations and analytics. We are fortunate that when we get there we will have David’s work and David’s example—neither of them small—to guide us.

University of Maryland
Works Cited

Garfinkel, Simson L., et al. 2006. “Advanced Forensic Format: An Open, Extensible Format for Disk Imaging”. In Advances in Digital Forensics II: FIP International Conference on Digital Forensics, ed. Martin Olivier and Sujeet Shenoi. 17–31. New York: Springer.

Greetham, David. 1996. “Textual Forensics”. PMLA 111.1: 32–51.

Howse, Martin. 2013. Diff in June. Brescia: LINK Editions.

Kruse II, Warren G. and Jay G. Heiser. 2001. Computer Forensics: Incident Response Essentials. Boston: Addison-Wesley.

McNeil, Heather and Bonnie Mak. 2007. “Constructions of Authenticity”. Library Trends 56.1: 26–52.