Data and Wisdom: Electronic Editing and the Quantification of Knowledge

Julia Flanders
Brown University, USA

Abstract

The concept of data in the humanistic academy carries a heavy cultural freight: as a reductionist yet efficient representation of complex textual significance. Far from being an invention of the digital age, this conception of the role of quantification has a prehistory whose terms continue to resonate in modern debates about digital editing and digitally mediated scholarship. This essay explores these terms and the anxieties they reflect, concluding that digital representation is no less textually and methodologically rich, and no less a production of knowledge, than its print counterpart.

Correspondence:
Julia Flanders, Women Writers Project, Box 1841, Brown University, Providence, RI 02912, USA.
E-mail:
Julia_Flanders@brown.edu

This essay was written in 1997. It has not been significantly updated but reflects that particular moment in the debate on electronic text.

1 Introductory

When pundits discuss the impact of computers on the history of the book, they inevitably come to a point where they feel compelled to say 'Of course, the book won't be disappearing any time soon ....' But in reassuring themselves about the physical object, they overlook a more profound change. Italo Calvino hints playfully at what might be happening to readers and texts in the age of computing in *If on a Winter's Night a Traveller*, where he presents a novelist pre-occupied with the ways his novels are read. One day, he says, 'I asked Lotaria if she has already read some books of mine that I lent her. She said no, because here she doesn't have a computer at her disposal.'[1]

But it gets worse:

> She explained to me that a suitably programmed computer can read a novel in a few minutes and record the list of all the words contained in the text, in order of frequency .... The idea that Lotaria reads my books in this way creates some problems for me. Now, every time I write a word, I see it spun around by the electronic brain, ranked according to its frequency, next to other words whose identity I cannot know... Perhaps instead of a book I could write lists of words, in alphabetical order, an avalanche of isolated words which expresses that truth I still do not know, and from which the computer, reversing its program, could construct the book, my book. (p. 188)

This is poignant enough, especially that final 'my book'; but there is still one further sorrow reserved for the novelist. With a novel stored in the computer, and about to be printed out, disaster strikes:

> at some point she must have pressed the wrong key. The order of the words in the text of Calixto Bandera [the novelist], preserved in the electronic memory to be brought again to light at any moment, has been erased in an instant demagnetization of the circuits. The multicolored wires now grind out the

Literary and Linguistic Computing, Vol. 24, No. 1, 2009.
doi:10.1093/llc/fqn036 Advance Access Published on 25 November 2008

> dust of dissolved words: the the the, of of of of, from from from from, that that that that, in columns according to their respective frequency. The book has been crumbled, dissolved, can no longer be recomposed, like a sand dune blown away by the wind. (p. 220)

I am not certain whether Calvino is here elegizing the novel, or poking fun at the ontological illusions of its creators, or directing our fascination to a deep-rooted conundrum. However, both the fictional novelist's dismay and Lotaria's briskly efficient techno-progressivism are positions which we see rehearsed more directly all around us these days: strong emotions attach themselves to either side of what has come to be 'the electronic text debate'. And where they focus most strongly, most subtly and (I think) most interestingly is in this matter of the decomposition of the text: a process which can be seen as leading on the one hand to a newly efficient purchase on the text, and on the other to a complete loss of everything for which the text is held to be valuable.

These two possibilities—grasp versus disintegration—seem to pull in opposite directions, but they are premised on a common sense of the profound difference between textuality and data. Both assume that the condition of data is one of quantification: a state in which information is subjected to the strict codification of the measurable, in which the terms of measurement are stated explicitly and in which a further numerical, statistical, quantificatory activity upon the information is thereby made possible. And both likewise assume that text on its own is not like this: that the quintessence of the textual condition is precisely its indeterminacy, its plenitude of intimations, its subtlety and implicitness. The difference lies in the value placed on each side of the balance. To someone in a text-loving hat, quantification looks like mere counting (always 'mere'), whereas under the data-loving hat, subtlety and intimations look like fog. The battle for control over the metaphor—whether we shall describe the polarities here as counting versus wisdom, or rigorous data versus fog—has its roots in some very basic cultural paradigms, but I would like to focus my discussion somewhat by considering its history specifically in connection with the intellectual and academic field of textuality and its study and management: what we call textual editing and literary studies. It is by the effort to define these disciplines that their practitioners found themselves structuring the intellectual terrain in this manner—not, certainly, inventing these polarities, but building them into the foundation of a new discipline as the condition for its operation and self-knowledge. In what follows, I am first going to sketch a contemporary framework for thinking about the nature of data from the humanist's point of view, and then amplify that characterization by reference to a particular historical debate, one in which we can see represented not only the emotion but also some of the underlying structural issues, which contribute to the way data, quantification and humanism are imagined today. Finally, I would like to situate our hopes and fears for electronic editing within the context thus established.

2 Humanism and the Nature of Data

If we consider the nature of data not in some absolute sense, but through the ways that the concept figures in the intellectual work of our culture, we can see immediately that it carries a heavy emotional and even ideological freight. This is particularly true of its role in discussions about the relationship between the humanities and the sciences, where the idea of data and its penumbra of issues (the possibility of objectivity, the value of statistical studies) becomes the crucial point of debate and self-identification. As a component of a statement about methodology, it can scarcely be deployed without implicitly stating the affiliation of the speaker, and not as a mere matter of fact but as a declaration of kinship, vested interest, antagonism, defensiveness and so forth. Its complex range of associations and values is extremely revealing, and it is worth considering some of these before we proceed, as a way of situating the concept in its cultural space.

Consider the following, extremely partial, definition:

> Data is information that scientists use in their work.

Depending on who is speaking, this could be the critical (or defensive) voice of a humanist, the affirming voice of a scientist or the aspiring voice of one who hopes to put humanistic study onto a newly empirical footing. Whichever the voice, though, it invokes the deep-rooted cultural assumption that what differentiates the sciences (for better or worse, or neither) from the humanities is the nature of their information (its exactness, quantifiability and claims to objectivity) and what they do with it (measurement, counting, statistical work, prediction). The rigor, or the reductionism, of this method are alike rooted in these qualities and regarded as constitutive of the scientific method (again, regardless of who is speaking).

Next, consider this second, more subtly partial definition:

Data is information that computers can read.

Here again, the voice could be approving or disapproving: this could be a measure either of the superiority of data or of its utter mereness. However, for the culture at large, and certainly for most humanists, it is likeliest to be the latter; it is the simplicity—the simplisticness—of the way computers 'think' that is taken as their most culturally important characteristic. The first book I ever read about computers (one written for children in the 1970s, mostly about ENIAC) thought that the key to understanding them was to know that they deal with binary data, ones and zeroes. Back when computers were a new thing in the public imagination we were told with considerable frequency that computers work only according to the most basic arithmetical and logical operations; that they are fundamentally 'stupid'. These statements are in some sense basically accurate, and yet I think we can also detect in them a tone of defensiveness and self-reassurance, as much as to say: computers would not replace humans, and even if they beat us at chess, they can only do it in a 'mechanical' way, not by the exercise of genius which is a uniquely human characteristic. They do it by counting, and they cannot even count higher than one. To say, then, that data is information which computers can read might be to emphasize its mereness with a vengeance, and to distinguish it from other, more human/humane/humanistic forms of information, forms whose complexity and richness make them impossible for the computer to process. Thus in the Calvino example, the novelist's fear fixates on the transformation of his text into data, and the way it comes to belong to the computer and thereby ceases to be 'his'.

These examples express in a somewhat vulgarized form the cultural and emotional associations which the concept of data has acquired. They echo, however, the more nuanced statements we hear in our own disciplines, particularly at moments where disciplinary boundaries are being approached or tested. This is particularly true of the boundaries between the humanities and the sciences, but even between different humanities disciplines (say history and literary studies), arguments will tend to coalesce with special frequency and energy around issues of the nature of the information studied, how and whether it can be measured or pinned down factually and whether this is a good thing. In computer-assisted literary study and scholarship involving electronic texts, however, these issues are most acutely felt, and the need to draw boundaries seems most urgent. Thus in an early discussion of computer-aided literary analysis taken from Roseanne Potter's *Literary Computing and Literary Criticism*, we see first an approving description of the computer as counting machine:

> Verification, though not a concept new to literary criticism, certainly represents a shift in focus away from brilliance of insight and assertion toward the detailed testing of scientific experimentation. . . . Objective treatments of texts frequently involve not only finding examples of features, but also counting them and comparing the results with known facts about language. Things counted produce sums; the existence of sums encourages comparison with other sums; statistical analysis follows almost inevitably.[2]

And then, immediately following, there is a sudden recoil:

> Only the presence of critical judgement saves the research from veering off into number juggling. . . . The usual impact of numbers on

> texts is reductionist. All the beautiful specificity of figures of speech can get lost when each detail is represented by a number. A balance must be carefully maintained between acquired scientific methods and critical values. (p. xvii)

This is worth unpacking: what, for instance, is the import of the statement that 'the usual impact of numbers on texts is reductionist'? What is meant here by 'representation'—in what sense are details being represented by numbers? on what criteria are we expected to value statistical analysis, verification and objective treatment of texts, and yet distinguish these from 'number juggling'? What is the nature of the 'balance' that is sought here? Above all, what is the motive, the prick of the humanistic conscience, that so quickly propels us to the words of caution at the end—especially given that it does not result in concrete specifications for avoiding number juggling or maintaining critical values, but leaves us only with a vague sense of dread?

To answer these questions, we need to focus more carefully on the nature of representation. When thinking of representation, and particularly the form of representation we call data, it is tempting to imagine a Thing being represented to which the representation offers us only partial access, but which we believe we might perceive better if we could only get close enough. This temptation is reinforced by the fact that when we create an electronic text, we are quite likely to be reproducing some physical text which we have in front of us, which acts as the Thing in a number of ways; it is causally and temporally prior, it is the authority in cases of disagreement, it is perhaps the thing towards which we feel more affection. In the case of electronic versions of a rare book or manuscript, the sensation of Thingness in the original is almost overwhelmingly reinforced by its cultural and monetary value, and by the fact that for scholars it is an object of specific study precisely by virtue of its physical format. We study book bindings, we do not study binhexing—although John Lavagnino has proposed the intriguing possibility of an analytical bibliography of electronic texts.[3] These considerations lead us to regard the qualities which this object does not naturally share with our new version—its physical and chemical properties, its visual appearance, its age—as the characteristics which really constitute its Thingness and its resulting claim on our interest, and we thus feel ironically impelled to consider our reproduction more accurate, more valuable, as it manages to remind us of these characteristics or (more frequently) to consider our reproduction to be the more obviously *not* the Real Thing to the degree that it fails to do so. Our reproduction, in other words, is *only* a representation.

Where these pre-occupations go astray is partly in their misapprehension of the nature of representation, but only partly. We all probably assent to the idea that our access to the real is irredeemably mediated by representation. We are not actively thinking that the Book is real and the electronic text is not. But we might be thinking that whether or not the Book is real, it is what we are paid to study, are interested in studying, are trained to study and that it does no good to say that we should really be interested in the electronic text instead. I suggest, then, that where we have strayed in this case is in assuming that the best way to study the book is by looking at it, the supposed Thing, as closely as we can. In fact, I would like to suggest, if we can represent it powerfully and usefully enough, we might be able to study it in ways that are not otherwise possible.

The real challenge here is understanding that some representations are better than others— in fact, better even than the ones we think of as the Thing Itself—in the sense of providing a more useful analytical model. Once we understand the purpose for which we need such a model, we will find that some models are far more powerful than others, and give us better leverage over the problems we are trying to solve. This does not necessarily mean that they are more detailed; on the contrary, for some purposes a simpler model is essential, just as a road map showing only the interstates is what you may need for driving cross-country, where a detailed topographical map with a scale of 1 in. to the mile (or one of those stiff contour maps with three-dimensional hills) would be hopelessly unwieldy. The reflexive correlation of detail with accuracy is a chimera produced by the idea that

the representation is trying, in a doomed and hapless way, to *be* the Thing. The goal of analytical representation should be rather to *model* the Thing in analytical ways. Indeed, the most important area of research in text encoding, I think, is the attempt to discover ways to represent all aspects of the text, including its physical and visual forms, in such a way as to give us model of the thing that functions, analytically, better than the thing itself.

Where computers are concerned, this model will of course be constructed, at some very low level of abstraction, out of numbers: ones and zeroes, if we are thinking of binary; a few more if we are thinking of machine language. But these days, the emphasis of encoding methodology has almost nothing to do with these numbers; the closest we come is in thinking about character sets. Our text is no more being 'represented by' a number than our thoughts are 'represented by' the reductive system of twenty-six characters that we call the alphabet. Data is thus becoming for us less of a purely numerative concept, and expresses something closer to the application of an interpretive or analytical strategy. 'Good data' from a text-encoding standpoint is data whose encoding scheme has been carefully constructed to identify textual features unambiguously, to group them into categories which express useful logical distinctions, and to express relationships between features in a way that usefully models the structure of the text. In text encoding, then, the quantification is of a different sort: it involves defining a closed set of terms by which to describe the text's structure and behaviour, and applying them consistently. The reductionism involved in this activity is thus only the reductionism of our descriptive model; it has nothing to do with the tokens, numerical or otherwise, which the computer uses internally when talking to itself. And again, as we saw above, the reductionism of a descriptive model need not represent for us an essential loss of meaning. On the contrary, if we have any experience thinking about designing text encoding systems, we will understand the loss as strategic, as a way of representing the text according to the qualities we find important to its analysis. Obviously, some things are easier to represent than others—no one has yet, to my knowledge, found a way to add anything to the study of metaphor by the use of text encoding—but actual *loss* will only occur where we attempt to represent something without a well-conceived scheme for doing so.

We have begun by considering data in some detail, and before moving on to some of the ancestry of this issue, I should mention the other term from my title, 'wisdom'. If data has been positioned here as a form of information which emphasizes the leverage we gain by creating structured, quantifiable models of the world, then wisdom as an approach to the universe on the contrary emphasizes what we cannot think through models or systems: it emphasizes those things which are not susceptible to data-like treatment or analysis. I have chosen this word because its positive connotations indicate the value we attach to the idea of holistic knowledge and judgement, and also because it seems to represent the quintessence of what we feel a computer can never be. The formulation 'data or wisdom' here is thus intended to bring to the fore a polarity of values, each of which seems attractive in its own right, but each of which is valued by its antagonism to the other. This antagonism forces us to ask our questions in a certain format: one which imagines that the answers we get from creating and working with structured models will always be those for which society pays top dollar and by which it gets to the moon, and yet will also always be reductive and always lack the qualities for which we value our humanity. In imagining the relationship between these two concepts, we can stand them up against one another as seductive but mutually irreconcilable poles between which we try to find a position of 'balance': rigorous statistical study tempered with critical values. But the very familiarity and naturalness of such an polarity should make us wary, should make us ask why it holds such attractions for us, and how those attractions are rooted in our institutional and disciplinary history. To see these kinds of allegiance cast in another form—to see them articulated through motivations and anxieties which echo our own only as our ancestors' faces, seen in dim daguerreotype, foretell our own physiognomy—is perhaps to see more clearly the institutional and intellectual framework which helps to

bring these allegiances into being and give them their force.

3 Editing and Quantification

We could go back quite far in investigating the roots of the editing profession, and its institutional history, but in the interests of brevity I am going to take quite a specific episode from the last century, namely the altercation which took place between Algernon Swinburne, the poet and literary critic, and Frederick Furnivall and other members of his New Shakspere Society [sic].

First, a bit of context. The New Shakspere Society was founded by Furnivall (who had also founded a number of other societies, notably the Chaucer Society and the Early English Text Society) in 1873 as a way of encouraging the study of Shakespeare and of getting reliable editions of his work published. Interest in Shakespeare at this time was considerable; the foundational premises of the Society, however, gave it an idiosyncratic turn which set the stage for its notoriety and for the embroilments in which it almost immediately found itself. The following quote from Furnivall's opening speech at the Society's first meeting gives a sense of its intellectual commitments:

> The purpose of our Society . . . is, by a very close study of the metrical and phraseological peculiarities of Shakspere, to get his plays as nearly as possible into the order in which he wrote them; to check that order by the higher tests of imaginative power . . . and then to use that revised order for the purpose of studying the progress and meaning of Shakspere's mind . . . [4]

The Society's work in fact began with several close studies of Shakespeare's meter, with a strong emphasis on attempts to use metrical tests as a method of dating the plays, and with considerable internal debate over the nature of metrical and other quantitative tests, and their role in literary study. Frederick Gard Fleay was one of the strong proponents—considerably stronger than Furnivall—of metrical tests, and of their value in setting literary study on what he thought of as a scientific basis. Thus in one of his earliest papers for the Society he argues:

> This, however, is the great step we have to take; our analysis, which has hitherto been qualitative, must become quantitative; we must cease to be empirical, and become scientific: in criticism as in other matters, the test that decides between science and empiricism is this: 'Can you say, not only of what kind, but how much? If you cannot weigh, measure, number your results, however you may be convinced yourself, you must not hope to convince others, or claim the position of an investigator; you are merely a guesser, a propounder of hypotheses.[5]

Fleay not only registers this difference of approach as marking the boundary between two heterogeneous undertakings, disciplines which are characterized precisely by the kind of work they do, but also wants to argue for making one more like the other: literary work, however, little it may seem to be concerned with scientific values, *ought* to be concerned with them:

> It may seem [he says] to some ludicrous to speak even of the application of mathematics to such a subject; but it will be seen from the table that the plays assigned to the period - exactly agree with those in Meres's list . . . Now, the doctrine of chances gives us as the odds against these 10 plays being selected out of the 30 . . . more than 20 millions to one . . . To the mind accustomed to the exact sciences, this fact alone is conclusive as to the immense value of the rhyme test. (p. 14)

The 'mind accustomed to the exact sciences', he implies, is one which is better qualified not only for science but also for the study of letters, as that study truly ought to be carried out.

The fracas in which the New Shakspere Society found itself arose from a profound disagreement on this subject, bringing into conflict not just Furnivall's and Swinburne's personal pugnacity, but two conceptions of how literary criticism can

arrive at just conclusions about its object of study. On 1 April 1876, after several articles by Fleay had appeared in literary journals propounding the 'quantitative criticism' of the New Shakspere Society, Swinburne published a parody in *The Examiner* entitled 'Report of the First Anniversary Meeting of the Newest Shakespeare Society' which presented the quantitative approach in an extremely ludicrous light. Its emphasis on tabulation of statistical results and on precise periodization of Shakespeare's plays becomes in Swinburne's hands an absurdly elaborated scheme of proof, animated by an overbearing self-assurance which is blind to the aesthetic nature of the plays:

> It was evident that the story of Othello and Desdemona was originally quite distinct from that part of the play in which Iago was a leading figure. This he was prepared to show at some length by means of the weak-ending test, the light-ending test, the double-ending test, the triple-ending test, the heavy-monosyllabic-eleventh-syllable of the double-ending test . . . [6]

Swinburne's parody associates this kind of statistically based research with a literalism which not only cannot comprehend the nature of metaphor and figuration, but also regards them as antique curiosities, rendered obsolete by the march of progress. Thus:

> Mr. D. then brought forward a subject of singular interest and importance—'The lameness of Shakespeare—was it moral or physical?' He would not insult their intelligence by dwelling on the absurd and exploded hypothesis that this expression was allegorical, but would at once assume that the infirmity in question was physical. Then arose the question—'In which leg?' (p. 381)

What Swinburne's characterization of the New Shakspere Society's membership emphasizes above all, though, is the disintegrative and reductive effects of their work: the way in which their quantifying approach, as he sees it, fails to represent the qualities that make the plays art. His remonstrance, expressed more fully in his introduction to *A Study of Shakespeare*, relies on a dichotomy which opposes 'the music which will not be dissected or defined'[7] to the 'purely arithmetical process' of 'counting up of numbers and casting up of figures' (p. 5), and it opposes 'the singer', the sympathetic fellow artist, to the 'pedant' and the 'sciolast' (p. 5) whose 'horny eye and . . . callous finger' are just barely sensitive enough to perform the basic counting required by their method. Numbers here are shown as radically incommensurable with a representation of the kinds of patterning which make a play a work of art: numbers are merely enumerable, where the 'music' of poetry contains 'infinite varieties of measure' (p. 5), 'delicate and infinite subtleties' which no amount of effort can count or tabulate.

By representing the argument in terms which set 'poetry' against 'numbers', though, Swinburne has already presupposed his conclusion, just as Fleay presupposes his when setting the 'investigator' against the 'guesser', or Furnivall his when putting the one who can 'weigh, measure, number' his results against the mere 'propounder of hypotheses' based on 'mere subjective feeling' (*NSST*, p. 19). The opposed quantities, 'poetry' and 'science', with all their attendant associations, had already by this time become entrenched not just as values in their own right, but as twin anchors for a world view in which they can be expected to play out a familiar drama of opposition, over and over, and thereby substantiate a larger paradigm to which both assent. This larger paradigm is in effect the *via media* which claims that science and poetry are both important: that we must work carefully to gather facts and observations, but then must exercise a wise critical judgement which can only be attained by intuition and experience. There is no space here to do more than just point out how this paradigm operates powerfully throughout the ideology of English culture; the notion of a fruitful cooperation between science and intuition, a marriage between manly rigor and feminine imagination, underpins social structures as various as aesthetics, marriage, politics and scientific inquiry, and it is one to which I think we still find ourselves recurring. However, further elaboration on this is a topic for another time.

4 Electronic Editing

Editing, at bottom, has always been a way of modelling a text or a group of texts so as to make them available for certain kinds of analysis. If, after the foregoing, this makes it sound a lot like text encoding, I think that is no mistake: certainly it has been observed before this that text encoding is a form of editing. However, editing has conceptualized the modelling it does in very specific ways, which focus on the relationships between different documentary states of the text and on indeterminacies within the text, such as errors, illegibility and unclear meaning. What I have tried to show thus far has been that we cannot afford to think of data, of analytical modelling, as something alien to our discipline: that in fact our positioning of it as Other stems more from a cultural paradigm rooted in the way our own discipline created itself than from anything about data itself, which is really about representation—a subject with which our discipline has everything to do.

For traditional editors, these issues could seem peripheral; for editors in the electronic age, they insinuate themselves into our work and our conferences, into the very centre of our attention—the more so because these issues are the focus of so much emotional energy, anxiety and enthusiasm alike. I have thus laid such elaborate groundwork because it seems to me to be the only way to gain self-consciousness about the questions that really concern us. I am now going to turn to these questions, the ones which have motivated the conference at which this article was originally presented: how do we think about the edition in the electronic age? Will it, or should it, resemble the traditional edition? How can we construct editions that will help us in our work and perform the cultural functions we value?

It is clear that we could use the electronic medium, as a number of people already have, to create traditional-style editions which are simply easier to navigate and harder to take to the beach. I think, though, that the conceptual issues posed by these editions have less to do with quantification than with editorial theory proper, and so I am going to set them aside to focus instead on the electronic editing projects, which are undertaking something entirely new, something which does not already have a precedent in the traditional editing world. For the sake of argument, I would like to propose that the kind of edition which best models the text for our scholarly use is something resembling an electronic archive, but one in which analytical and editorial relationships between texts are represented by encoding and by computational relationships. Such an archive would offer the primary sources essential to further scholarship, tools which allow these sources to be studied separately and in combination, and structural modelling to enable textual analysis and literary study. As an example, just to prove that this is not some sort of addled wish fulfilment, I adduce the Canterbury Tales Project undertaken by Peter Robinson.[8]

As editors, how are we to understand the work that creates this resource? Peter Robinson, who has perhaps gone the furthest in this kind of work, has articulated a certain perplexity about this question:

> When we publish all this, what are we going to call it? You could call it an archive, a dossier, a resource base. But is it an edition? If an edition is something you pick up and say 'Now I have it, the text as Chaucer wrote it', then it is not an edition. I began my scholarly life as what you might call a proper editor . . . This is what I thought editors did—they presided over the text: they read, they weighed evidence, and they decided. But now, I am no such editor. I am a software compiler and developer . . . a manuscript entrepreneur; worst fate of all, a transcriber.[9]

Kathryn Sutherland has similarly asked us to consider what might be the cost of reconceiving the electronic edition as an electronic archive, and suggests that the newer term represents to some degree an abnegation of authority on the part of the editor, either from a desire to dissipate the locus of that authority altogether, or from a desire to relocate it onto the reader.[10] She suggests as well, I think, that the evacuation of this authority may leave a gap which we will no longer be capable of filling. As she says,

> Herein lies the greatest challenge posed by the electronic environment . . . If the computer merely displays knowledge to a post-productive society, what might this imply about our mechanisms for generating new (as opposed to retrieving and redeploying old) expert knowledge? How real is the danger that the scholar-worker, whose origins lie in a nineteenth-century conception of learning as heroic endeavour, will be transformed into the scholar-technician? (p. 18)

Or, as she puts it in an earlier essay, 'Do [computers], in taking the labour and randomness out of intellectual inquiry, remove much of the knowledge, too?'[11]

Whether or not we think the answer is 'yes', the logic of these questions is worth unpacking, because it gets at the heart of our own stake in how electronic editions represent texts, in how data and text interact, in where wisdom may fit in. In Sutherland's phrasing, the 'scholar-technician' is not a producer but a reproducer of knowledge,[12] precisely to the degree that he or she is no longer involved in the weighing and deciding of which Robinson speaks. But to take the opposite position for a moment: to say that the scholar–technician is not producing knowledge is not to say that no one is producing knowledge, or that the use to which primary texts used to be put will no longer exist. To counter the vision of a knowledge-free world, we have only to insist that the difference between the new world and the old will be one of access and division of labour, not of end product. The resources we regard as the front line of available information—the primary texts we consult—will be truly primary sources, made available in a form that lets us exploit their primariness, and the work of producing knowledge based on those sources—literary study, the creation of editions—will continue to be done by scholar-workers, and also by teacher–workers, and possibly student–workers as well. If the concern then becomes a fear that the knowledge thus produced will be of an inferior quality—and perhaps that we would not even be aware of its inferiority—this is a concern of a different sort, and with different remedies. It is in fact a concern over our own cultural importance as editors, and over the importance of the work we do.

To explain the case in these terms is to suggest that the creation of these primary text resources—'archives' or whatever we want to call them—is not editing, and that editing will continue to go on in a realm apart, according to its established usages. In fact, however, what is happening is that the very idea of the edition is being offered avenues of expansion which might redefine our work—redefine it not so as to embrace joyously the idea of being (merely) a 'scholar-technician', but so as to make us see these new archives as constituting the production of knowledge. Peter Robinson does, after all, conclude that he is an 'editor', and that the Canterbury Tales project will result in an 'edition' (p. 9), albeit one which involves 'redefining the relationship of an editor to what he or she is editing' (p. 10). If I may expand on what Robinson explicitly proposes, to suggest what I think is groundbreaking about his work, I think that this redefinition above all can be seen in the transformation of relationships between texts from being commentary to being analytical or even computational linkages, and from being detail to being part of a representational model. Thus, taking the Canterbury Tales project as an example, the relations of variance between one text and another—variations of spelling, of letterform, of word order, etc.—are expressed by the text's encoding so that they can represent a full model of the process of transfiguration which takes us from one version of a text to any other. By 'a full model' I mean here something profoundly different from a list we can read in the back of the book, something closer to an algorithm, a computational expression of the relationships between text and text which can enable us to study those relationships with a power we did not have before. The 'edition' of the Canterbury Tales thus produced is, as Peter Robinson says, 'an edition of all the text in all of the . . . witnesses' (p. 10), but it is also an edition of all the texts which those witnesses collectively produce: not just an archive, as it would be if each transcription were isolated from all of the others, but an edition which gives us analytical leverage that we could not otherwise attain.

In order to get this analytical leverage, though, we need to accept the fact that the edition is a representation, and that it is precisely its representedness that enables us to gain our purchase on the text. This acceptance involves sacrificing the illusion that we are dealing with the thing itself, or that we should be; we need to understand that representation is the textual condition, not to be evaded by any means. I would like to suggest as well that we may work more comfortably with electronic resources if we stop thinking of the quantitative realm as the realm of the reductive, but rather as the realm of representation: the realm in which we can create analytical models of text through which we can gain new critical purchase on them. The most powerful editions of the electronic era will be those that use this kind of modelling to the fullest: using text encoding to create models of the text's structure, using those models to morph the text, to reorder the words, align them with other versions, lemmatize them and so forth. Some will say that such a thing does not look very much like an edition any longer. This may in fact be so: perhaps 'edition' would not be the word we want to use for these things. However, they will fill the cultural space of an edition, and perform what are currently its most useful functions: to provide access to the texts of the past in a form that we can use. I think we are seeing that the idea of 'use' is changing considerably, but I think we should fear these changes less than we should fear disuse. If these changing expectations and capacities issue a challenge to editors to reconceive their work, then as editors I think the only thing to do is accept the challenge—not to struggle to make our same 'editions' in a new medium, but to make texts for use on which wisdom is worth expending.

Notes

1 **Calvino, I.** (1979). In Weaver, W. (trans.). *If on a Winter's Night a Traveler*. New York: Harcourt Brace Jovanovich, p. 186.

2 **Potter, R. G.** (ed.) (1989). Preface. Literary Computing and Literary Criticism: Theoretical and Practical Essays on Theme and Rhetoric. Philadelphia: University of Pennsylvania Press, p. xvii.

3 **Lavagnino, J.** (1996). 'The Analytical Bibliography of Electronic Texts', a paper presented at the joint annual conference of the Association for Literary and Linguistic Computing and the Association for Computers and the Humanities, Bergen, Norway.

4 **Furnivall, F. J.** (1874). Director's Opening Speech. In *The New Shakspere Society's Transactions*, Vol. 1. London: Trübner and Co., p. vi. (Cited hereafter as *NSST*.)

5 **Fleay, F. G.** On metrical tests as applied to dramatic poetry, in *NSST*, p. 2.

6 **Swinburne, A. S.** (1876). Report of the first anniversary meeting of the Newest Shakespeare Society. *The Examiner*: 381–2. This article was reprinted with slight alterations as an appendix to Swinburne's *A Study of Shakespeare* in 1880.

7 **Swinburne, A. S.** (1880). *A Study of Shakespeare*. London: Chatto and Windus, p. 6.

8 In the 10 years since this article was originally presented, this approach has been adopted more widely; furthermore, digital publication tools (for instance, the Anastasia software developed by Peter Robinson for Scholarly Digital Editions) have been developed which support this approach and take its desirability for granted.

9 **Robinson, P.** (1993). Manuscript Politics. In Warren, C., Caroline, D. and Marilyn, D. (eds), *The Politics of the Electronic Text*. Oxford: Office for Humanities Communication Publications 3, p. 9.

10 **Sutherland, K.** (1996). Looking and Knowing: Textual Encounters of a Postponed Kind. In Warren, C., Marilyn, D. and Andrew, G. (eds), *Beyond the Book: Theory, Culture, and the Politics of Cyberspace*. Oxford: Office for Humanities Communication Publications 7, p. 15.

11 **Sutherland, K.** (1993). Challenging Assumptions: Women Writers and New Technology. In Warren, C., Caroline, D. and Marilyn, D. (eds), *The Politics of the Electronic Text*. Oxford: Office for Humanities Communication Publications 3, p. 65.

12 The gender implications of this distinction are reinforced in Sutherland's argument by a discussion of the concept of the 'virtual', in which 'the computer's transformation of the labour of learning into the seductive but less virtuous act ("virtue", "the possession or display of manly qualities", OED, sense 6) of merely looking [might] come to denote the same feminization of endeavour (by which I mean cultural disempowerment) it traditionally has done' (Sutherland, 'Looking and Knowing', p. 19).