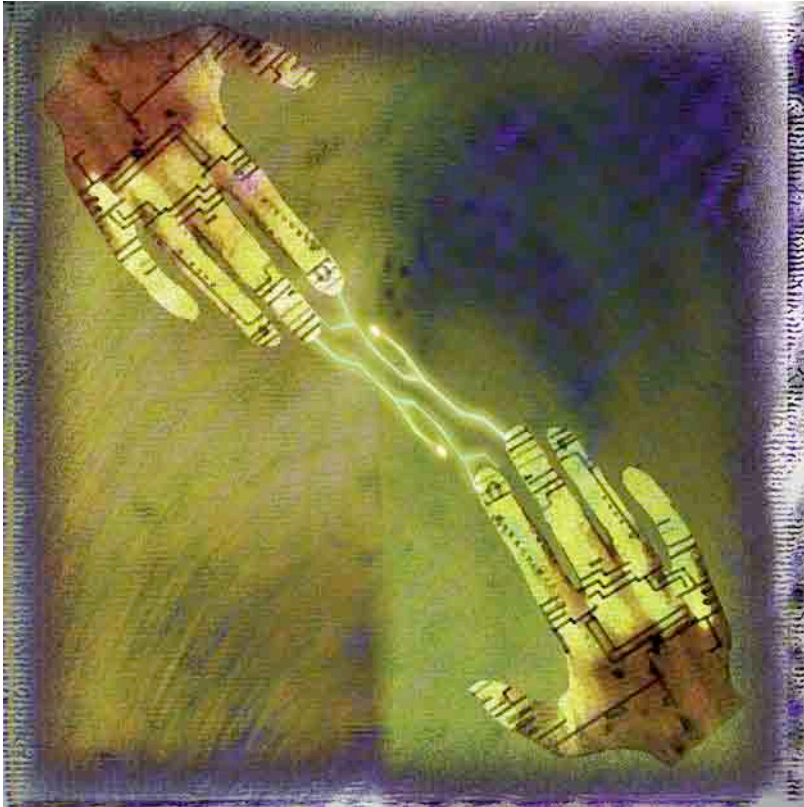


# LYMAN AWARD

## Digitizing the Humanities: The Richard W. Lyman Award Lectures

ROBERT K. ENGLUND  
WILLARD MCCARTY  
JEROME MCGANN  
ROY ROSENZWEIG  
JOHN M. UNSWORTH



**The Richard W. Lyman Award**

Presented each year from 2002 to 2006, the Richard W. Lyman Award recognized scholars who advanced humanistic scholarship and teaching through the innovative use of information technology.

To be considered for the award, individuals and teams presented work that created new knowledge in some domain of the humanities; that embodied technological innovation with broad application in scholarship and teaching; that addressed social, cultural, and/or economic issues in the creation and dissemination of scholarly work in the contemporary world; and/or work that used technology in new ways to bring the results of humanistic scholarship to student and public audiences.

Named in honor of Richard W. Lyman, president of Stanford University from 1970 to 1980 and of the Rockefeller Foundation from 1980 to 1988, the award was made possible through the generosity of the Rockefeller Foundation.

# LYMAN AWARD

Since man is a child of God and technology is a child of man, I think that God regards technology as a grandfather regards his grandchildren.

— **Father Roberto Busa,**  
*the first scholar to apply computers  
to the study of literature*



© 2007 National Humanities Center  
ISBN 978-0-615-21570-9

Photographs: Ron Jautz and Peggy VaGenius

National Humanities Center  
7 T. W. Alexander Drive  
P.O. Box 12256  
Research Triangle Park, NC 27709  
nationalhumanitiescenter.org  
(919) 549-0661



# CONTENTS

## Foreword

James O'Donnell ..... i

## Textonics\*: Literary and Cultural Studies in a Quantum World

Jerome McGann ..... 1

## Promises and Perils of Digital History

Roy Rosenzweig ..... 13

## Would We Have Noticed the Loss of the Iraq Museum?: The Case for Virtual Duplicates of Cultural Heritage Collections

Robert K. Englund ..... 23

## New Methods for Humanities Research

John Unsworth ..... 33

## The Imaginations of Computing

Willard McCarty ..... 47



# Foreword

Looking back at the future is a disorienting experience. The Richard W. Lyman Award gave us an opportunity in the first years of the new century to take stock of a futuristic revolution and recognize its leaders. The publication of this volume of essays lets us turn again to look ahead once more.

For one future is most emphatically over. Every fall Beloit College publishes a teasing survey designed to remind academics just how much the mental universe of the freshest eighteen-year-olds differs from that of their teachers. All of us were sobered to realize that this year's first-years had never been alive under a president not named Bush or Clinton, but a provostial examination of institutional records shows that it is not only our students who turn over at an astonishing rate. On my campus, fully two-thirds of the faculty and staff have come to us since Bill Gates introduced Windows 95 to a breathless world and Jeff Bezos sold his first book. For all that our institutions seem ancient, hoary, and traditional, under the thrall of graybeards, they are in

fact great whirling machines of human intelligence and inventiveness, constantly renewing themselves, constantly finding new futures.

The future this book captures is the future that hove into view in the late 1980s and early 1990s when the paradigm of the desktop computer took over the lives of prosperous Western societies. We could see the power of the devices themselves (though characteristically we did not well estimate just how much more powerful the things under our fingertips would become) and some at least could imagine what would happen when a worldwide network of such machines began talking to each other. That paradigm is now mature and even the developing world has been set free—or put in thrall—by its power.

The paradigm of the networked desktop computer emerged, moreover, in academic settings, a research project of engineers and scientists, shared with their humanistic colleagues, and for a long time almost confined to our



*l to r:  
James O' Donnell,  
Roy Rosenzweig,  
Jerome McGann,  
Willard McCarty,  
Robert Englund, and  
Geoffrey Harpham,  
President and Director  
of the National  
Humanities Center  
[Not pictured:  
2005 Lyman recipient  
John Unsworth]*

midst. It seems quaint now to remember the earnestness with which we said to each other in those days, “The Internet, you know, is off-limits to all commercial activity.” The clearest vision I now recall from those days was that of one of our awardees, Willard McCarty, who said to me at a meeting in Chicago in early 1992, “I think we’ve got about five years left”—before the commercial sector discovered, moved in, and took over the neighborhood. He was dead right, and a little optimistic.

Not for this book is to consider what happens now, when a new paradigm of handheld wireless devices emerges and dominates. The “One Laptop Per Child” project looks to put a \$100 laptop in the hands of as many people in emerging markets and developing nations around the world as possible, but the progress of that very academically based project is challenged at least a little by the realization that the target audience may be too busy text-messaging, Web-surfing, and chatting on something called a “phone” to be as interested in and desperate for what a “computer” can do as we imagined only a very few years ago. We should recognize, however, that the moment this book captures is definitely the moment—and only a moment—of “the computer age,” an age now ending.

The scholars recognized by the Lyman awards and appearing in this book, however, were never to be held back by fetishizing particular devices. They are the innovators who have theorized a more expansive future than computers or even phones can capture and who have accomplished important work animated by that vision. Well aware of the fragility of such constructions, we can call their imagination digital, in recognition that the fundamental underlying transition of our information age is the move from analogue (textured, powerful, noninteroperating) to digital (eminently interoperating, vastly powerful, but remarkably less textured and differentiated) representation of human knowledge and human cultural achievement. The essays here open windows into their vision and that future.



*Members of the Lyman selection committee at work*

The task of the Lyman Award selection committee, which I had the honor to chair for the five years of the award’s life, was to identify those scholars in humanistic disciplines who were doing and accomplishing substantial scholarly work that made use of, and would not have been possible without, digital information technology. It was never a technology award. We gave no points for using the latest, greatest, flashiest gadgetry. It was a labored but mandatory witticism every year for me, in introducing the award ceremony, to protest that we were interested in substance, not “bells and whistles” and that indeed we hoped for the Lyman Award to be thought of as a “no bells” prize. Our intention was to keep the focus on the work of the humanities and the possibilities for enriching human understanding and deepening the cultural and intellectual experience of scholars, students, and publics by the resourceful and thoughtful use of new tools. Implicit as well was the task of looking ahead to theorize what might become of us as knowers and creators of knowledge when the new tools had done their work reshaping the conditions of knowledge and the structures of society around us.

No such tasks are ever completed. In the five years of award-giving, we recognized and featured the work of creative, innovative, farseeing—and modest scholars. They are modest in the sense that they know that what we know now is provisional and everything we say about the subjects addressed here is subject to urgent revision—and very soon. The pages they have written here still stand up well on rereading, but like all important writing, these are pages that change constantly as the world that receives them changes, and

rereading them already allows opportunity for meditating on what is changing all around us as we read.

Four of our five recipients are scholars of textual culture above all else, one a historian. Jerome McGann, student of Romantic English literature of unparalleled eminence, has also been for twenty years a user and theorizer of the technologies of today. His Rossetti archive embodies and his *Radiant Textuality* (2001) springs from embodiment to imagination of the space in which we now move. John Unsworth, founding editor of the first online journal in the humanities (*Postmodern Culture* [1990–]), led the Institute for Advanced Technology in the Humanities at the University of Virginia for a decade and now is dean and professor in the Graduate School of Library and Information Science of the University of Illinois, where he continues his most important work as impresario and advance scout for transformations already under way. His lead work on the ACLS-sponsored report *Our Cultural Commonwealth* (2006) has set important principles in place for those who consider the “cyberinfrastructure” that the humanities require as a condition of continued innovation. Robert Englund, eminent Assyriologist and Sumerologist at UCLA, has remained focused on the most traditional of disciplines—recovering the most ancient textual past from cuneiform tablets, but his talk here reminds us how urgently vivid and present the cultural tasks of the humanities become in moments of crisis. Willard McCarty, holding the near-unexampled title of Professor of Humanities Computing at King’s College London, is a master theoretician of the present precisely because he is so unremittingly and deeply rooted in the ancient, medieval, and Renaissance pasts of our cultures. His work for twenty years now on building and hosting the global virtual salon of “Humanist” has made him pilot and mentor to two generations of innovators, students, and achievers.

The future escapes our grasp in more ways than one, and poignancy, indeed tragedy, can be the consequence. We have already lost one of our awardees to the grimmest of

fates. Roy Rosenzweig, social historian who became leader of an extraordinarily innovative Center for History and New Media (whose motto boasts of “building a better yesterday, bit by bit”) at George Mason University, lost a relentless battle with cancer in 2006. He left us just at a moment when recognition and opportunity were coming to the Center in the form of a major grant from the Mellon Foundation to take the Zotero tool the Center had created and make it the engine for a potentially transformative partnership with the Internet Archive to build ways of organizing and using vast open stores of cultural record. His loss reminds us again of my theme: futures constantly captured, new ones constantly in view, and the relentlessness of time and death in reminding us of the urgency and importance of such work.

I will not try to summarize or speak for the authors of the papers collected here. Few others are so skilled at doing so for themselves, even when, in Roy’s case, the voice comes now from beyond the grave. A few thanks are in order, however.

The Richard W. Lyman Award was funded for five years by the Rockefeller Foundation in honor of their former president, himself also the former president of Stanford University. All of us involved in the Lyman Award are grateful to the Foundation for its generosity in making the award possible



Richard W. Lyman

and to Dr. Lyman for graciously allowing us to use his name and for attending the first award ceremony in New York City in 2002. Lynne Szwaja, now of the Luce Foundation, was instrumental in working with us from Rockefeller.

The successive presidents of the National Humanities Center, W. Robert Connor (now of the Teagle Foundation) and Geoffrey Harpham, gave the award a home within an institution that deserves to be called what George Steiner thought was a dream, a true “house of reading,” a place where tranquility and attention make it possible each year for dozens of scholars to achieve work of lasting

importance. The vision of these presidents in seeing that reading itself is a work-in-progress and that scholarship changes as its media change, but in seeing also the importance of focusing on ends rather than means, on knowledge rather than information, has been essential in our work.

At the Center, Robert Wright and David Rice (both now of Duke University), and Joel Elliott were our facilitators, wire-pullers, and rescuers as we did the work on the selection and presentation of the awards. The selection committee grew and changed over five years, especially as we added each year's winner to the group, but it is a pleasure to think of the collegiality of these friends in the selection committee meetings that were more like seminars on the future of the humanities:

**Carla Antonaccio**

*Professor of Archeology and Classical Studies, Duke University*

**Peter Bardaglio**

*Interim Vice President and Dean for Academic Affairs/Professor of History, Goucher College*

**Consuelo W. Dutschke**

*Curator, Medieval and Renaissance Manuscripts, Columbia University*

**Robert K. Englund**

*Professor of Near Eastern Languages and Cultures, University of California, Los Angeles; Principal Investigator, Cuneiform Digital Library Initiative*

**Jerome McGann**

*John Stewart Bryan University Professor, University of Virginia*

**S. Georgia Nugent**

*President, Kenyon College*

**Roy Rosenzweig**

*Distinguished Professor of History and Cultural Studies and Director, Center for History and New Media, George Mason University*

**John M. Unsworth**

*Dean and Professor, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign*

*\*Affiliations reflect positions during tenure on the committee.*

Each year's selection was announced in a ceremony that was itself a gathering of eagles, three times in New York City (two of those at the New York Public Library), and once each at the Library of Congress and at the Newberry Library, all those libraries testifying to the importance and power of memory and the commitment of our culture to using memory to open up our understanding of present and future challenges.

Finally, I thank and salute an old friend, mentor, and inspiration, Professor Robert Hollander of Princeton University, sometime chair of the Board of Trustees of the Center, who chaired the advisory committee for the award and whose presence at the last of the award ceremonies was itself a testimony to the power of intellect in the face of temporality. His achievement in creating the digital Dante was exemplary on many levels and an inspiration to many who follow now at a distance, perhaps not even knowing who he was and who some of the other pioneers were who blazed these trails. McGann, Rosenzweig, Englund, Unsworth, and McCarty embody a generation of pioneers whose leadership stretches out into remote corners of the globe and will stretch into futures we do not now imagine. It has been a privilege of the first order to be associated with this act of recognition, memory, and imagination. The essays collected here well capture the luminous quality of this experience.

—James J. O'Donnell

*Provost, Georgetown University  
Chair, Lyman Award Selection Committee  
Former Trustee, National Humanities Center*

# Textonics\*: Literary and Cultural Studies in a Quantum World

Jerome McGann

*The following lecture was presented by Jerome McGann, the John Stewart Bryan University Professor at the University of Virginia, at the National Humanities Center on October 3, 2002. Professor McGann's digital/scholarly credentials include the Rossetti Archive, a hypertextual instrument designed to facilitate the study of Dante Gabriel Rossetti; the Ivanhoe Game, a Web-based software application for enhancing the critical study of traditional humanities materials; and extensive scholarly writings on computing in the humanities, including Radiant Textuality: Literature after the World Wide Web (Palgrave/St. Martin's, 2001). A noted scholar of the Romantic and Victorian poets and of textuality and traditional editing theory, McGann has also written several books of poetry.*



\*1. the semiological arts in general; especially the art of making things that have both beauty and use; 2. the semiology of virtual structures.

—Webster's New Virtual World Dictionary, revised

In play, there are two pleasures for your choosing,  
The one is winning, and the other, losing.

—Byron, *Don Juan*

He saw his education complete, and was sorry he ever began it. As a matter of taste, he greatly preferred his eighteenth-century education when God was a father and nature a mother, and all was for the best in a scientific universe. He repudiated all share in the world as it was to be and yet he could not detect where his responsibility began or ended.

—Henry Adams, *The Education of Henry Adams* (1907)

I.....

I'll come back to Byron later. Let me begin with Adams, whose urbane pessimism gets summarized in that late passage from his famous autobiography. An education ought to make one ready for life, but Adams's education has turned out to be a kind of black comedy. His humanistic training has left him unprepared for the dynamo of the twentieth century, which he is able to grasp only in its arresting superficialities—in its images, as he tells us in his

penultimate chapter—not in its gritty fundamentals. He *only sees* what is happening, he knows he has not *seized* it. So he joins the coming race as an observer, a scholar—what he calls a “historian.” But “all that the historian won was a vehement wish to escape.”

Today, as we pass through a similar historical emergency, a moment even more troubling for a humanist than Adams’s moment, *The Education* seems especially pertinent. We don’t want to guide our passage through this moment with tabloid reports like *The Gutenberg Elegies*, which supply us with a cartoon set of alternatives. Information technology comprises an axis of evil that Birkerts advises us to “refuse.” We can no more “refuse” this digital environment than we can “refuse” the empire our country has become. We may well feel “a violent wish to escape” both of these unfolding—and closely enfolded—histories, but we would do better to recall that we are characters in these events and so bear a responsibility toward them.

And there precisely we find Henry Adams waiting for us, caught between two worlds. Not between a dead world and a world powerless to be born, however, but between two living worlds, one relatively young, the other ancient. He neither abandons the one nor refuses the other. The positive revelation of his great book tells us that we all always inhabit such a condition. At certain historical moments that universal experience seems especially clear, and certain figures come forward to render an honest accounting.

The book also tells a cautionary tale, however, which is the second gift it passes on to us. If the dynamo and the Virgin each have their humanities in Adams’s view, he represents himself as the Nowhere Man. Not that he takes no action, but that he restricts his action to honest reporting. As a consequence, both Virgin and dynamo emerge from his book as mysterious forces—in fact, as those “images” that so preoccupy and immobilize him throughout his book.

I was asked to speak here today on the subject of “Where Will Information Technology Leave Humanities Education Five/Ten/Twenty/...N Years from Now?” The question implicitly asks for something more than an honest report. Reading Adams helps me remember what at my better moments I know: that I have little reason for confidence in my understanding, and least of all in any prognostic

powers. But he also reminds me that I do have hopes, as well as a few convictions about what we should look to be doing to shape those imagined futures ahead of us.

So let me begin with a conviction: that we have to carry out what Marxist scholars used to call “the praxis of theory”—or as the poet (Roethke) better said, we must learn by going where we have to go. Involved here are two hard sayings that can no longer be fudged or tabled. First, integrating digital technology into our scholarship will have to be pursued on as broad a scale as possible. Circumstances are such that this work can no longer be safely postponed. Second, we have to restore textual and bibliographical work to the center of what we do.

“What are you saying? Learn UNIX, hypermedia design, one or more programming languages, or textual markup and its discontents? Learn bibliography and the sociology of texts, ancient and modern textual theory, history of the book?” “Yes, that is exactly what I am saying.” And of course you ask why. At this point I give only one reason, though by itself—if we draw out its implications—the reason will more than suffice: because digitization is even now transforming the fundamental character of the library. The library, the chief locus of our cultural memory as well as our central symbol of that memory’s life and importance. That transformation is already altering the geography of scholarship, criticism, and educational method throughout the humanities and it forecasts even more dramatic changes ahead, as I shall indicate later. Moreover, the shifting plates are already registering on the seismographs.

Let’s begin at that point, with the signals coming from current, well-known events. First of all, some happy signs of the times. Already the library’s reference rooms are well along to virtually complete virtualization, and it’s difficult to believe any scholar regrets this. The transformation reflects the relative ease with which expository and informational materials translate into digital forms. To have immediately available to you those resources, wherever you might choose to set up your computer and go online, is a clear gain, and for older persons, an amazement. Such things can turn the soberest scholar into a digital groupie. Young persons tend to take such marvels for granted.

We want to cherish that generational difference when we begin to pick up on some other less happy signals. A grace of time is playing through the difficult period humanities education is now experiencing. But time takes time and some serious problems are short term, even immediate. A widespread malaise has been notable in our discipline for more than a decade at least, particularly among those heavily invested in humanities research education. One of the sources of this malaise—it has many—was addressed by a special letter sent to the members of the MLA last May by Stephen Greenblatt, the organization's president. Greenblatt pointed to publishing conditions that make it difficult or even impossible for young scholars to meet current standards for tenure in research departments of literature. He called the problem, correctly, a "systemic" one. A network of relations has bound together for a long time the work of scholarship, academic appointment, and paper-based—in particular, university press—publishing. This network has been breaking up, or down, for many years, and the pace of its unraveling has recently accelerated. In a grotesque inversion of our most basic goals, near-term economics, not long-term scholarship, has been a serious factor in humanities research for some time. Just *try* to find a publisher for primary documentary materials, or for any basic research that doesn't come labeled for immediate consumption: "Sell this by such and such a date"—before it spoils.

Do you see a digital savior waiting to descend? Do you think / see this redeemer? Well, I don't. But I think I do see that these broad institutional problems intersect with the emergence of digital technology, and that we won't usefully address the former unless we come to terms with the latter. The engagement won't solve our problems but it will help us to see them more clearly. Let me explain by recalling briefly a related part of our recent institutional history.

For as long as I've been an educator—since the mid-1960s—a system of apartheid has been in place in literary and cultural studies. On one hand we have editing, bibliography, and archival work, on the other theory and interpretation. I don't have to tell you which of these two classes of work has been regarded as menial if somehow also necessary. And like any system of apartheid, both groups were

corrupted by it. As Don McKenzie once remarked, material culture is never more grossly perceived than it is by theoreticians, whose ideas tend to remove them from base contacts with the physical objects that code and comprise material culture. But of course, as he went on to remark, the gross theoretician met his match in the myopic scholar, who gets lost in the forest by trancing on the bark of the trees.

To this day at my own university—an institution known for its commitment to serious work in textual and bibliographical studies—most of our advanced graduate students could not talk sensibly, least of all seriously or interestingly, on problems of editing and textuality and why those problems are fundamental to every kind of critical work in literary and cultural studies. I no longer ask our students in their PhD exams to talk about the editions they read and use, why they choose this one rather than another, what difference it would or might make. It goes without saying that these are bright and hardworking young people. Nonetheless, the institutional tradition they have inherited largely set those matters at the margin of attention, and never more unfortunately so than in the last quarter of the twentieth century. Until that time the American research program in English studies regularly made history of the language, editing, and bibliographical studies a requirement of work. I know from my own, painful experience that these requirements were often taught in killingly mindless ways, reinforcing our sense that they had nothing to teach us about literature, art, and culture—either of the past or the present. As we all know, in our country these requirements were universally dropped or eviscerated between about 1965 and 1990. (In England and Europe the situation is very different. Highly developed philological traditions permeate their scholarship.)

When I describe our recent educational history in these terms, I am sometimes suspected of fellow-traveling with a cadre of moralizers and educational instrumentalists. But remember, Bennett, Bloom, DeSousa, and Lynn Cheney are not enemies of theory or interpretation, they are simply strict constructionists in a field where Cornell West, Catherine Simpson, Edward Said, and Stanley Fish have been looking to broaden our ancient ideal of liberal education. Seeing the educational history of the past fifteen or

twenty years in terms of the celebrated struggles between these groups has obscured our view of an educational emergency now grown acute with the proliferation of digital technology. I'm no prophet and I hope I'm no Prufrock either. But there's a slow train coming and its song goes something like this: *In the next fifty years the entirety of our inherited archive of cultural works will have to be reedited within a network of digital storage, access, and dissemination. This system, which is already under development, is transnational and transcultural.*

Let's say this prophecy is true. Now ask yourself these questions: "Who is carrying out this work, who will do it, who should do it?" These turn into sobering queries when we reflect on the recent history of higher education in the United States. Just when we will be needing young people well trained in the histories of textual transmission and the theory and practice of scholarly method and editing, our universities are seriously unprepared to educate such persons. Electronic scholarship and editing necessarily draw their primary models from long-standing philological practices in language study, textual scholarship, and bibliography. As we know, these three core disciplines preserve but a ghostly presence in most of our PhD programs.

Designing and executing editorial and archival projects in digital forms are now taking place and will proliferate. Departments of literary study have perhaps the greatest stake in these momentous events, and yet they are—in this country—probably the least involved. The work is mostly being carried out by librarians and systems engineers. Many, perhaps most, of these people are smart, hardworking, and literate. Their digital skills and scholarship are often outstanding. Few know anything about theory of texts, and they too, like we literary and cultural types, have labored for years in intellectually underfunded conditions. It has been decades since library schools in this country taught courses in the history of the book. Does it shock you to learn that? We aren't shocked at our own instituted ignorance of history of the language or of bibliography.

Restoring intimate relations between literarians and librarians, a pressing current need, has thus been hampered by institutional developments on both sides. Insofar as

departments of literature participate in the work and conversations of digitized librarians, it happens through that small band of angels who continue to pursue serious editorial and bibliographical work: scholarly editors and bibliographers.

OK, then, what's the problem? Our traditional departments *have* managed to keep around a few old-fashioned editorial and bibliographical types. Let's send them out to help with the technical jobs and hope that their (that's *our*) brains aren't completely fried by beetle-browed and positivist habits. Once upon a time even *they* (that's *we*) were involved with the readerly text, right?

Those contacts might perhaps prove barely sufficient were it not for another recent upheaval in the world of higher education. For it happens that between about 1965 and 1985 textual scholars began to rethink some of the most basic ideas and methods of their discipline. I chose those dates because Ernest Honigman published *The Stability of Shakespeare's Text* in 1965, and in 1985 Don McKenzie delivered his famous inaugural Panizzi Lectures, *Bibliography and the Sociology of Texts*. So disconnected had the general scholarly community grown from its foundational subfield of textual and bibliographical studies, however, that this historic moment passed it by with little notice. The "genetic" and "social" editing theories and methods that emerged in those years signaled a major shift in literary and cultural scholarship. Because this change overlapped with the more public emergence of what would be called Literary Theory—perhaps "underlapped" is the better word—it drew scant attention to itself in that more visible orbit of literary and cultural studies.

A publication scheduled for later this year measures the change that overtook textual scholarship at the end of the last century. In 1982 Harold Jenkins published his celebrated edition of *Hamlet* in the Arden Shakespeare series. A lifetime's work, the book epitomized a traditional, so-called eclectic approach whereby Jenkins educed a single text of the play out of a careful study of the three chief documentary witnesses. At the end of this year a new Arden Shakespeare *Hamlet*, edited by Ann Thompson and Neil Taylor, will replace Jenkins's remarkable work. The new Arden *Hamlet* will not publish a single conflated text, it will

present all three witnesses—F1 (1623), Q1 (1603), and Q2 (1604–5)—each in their special integrity (or lack thereof).

The *New Yorker* magazine reported this event in a substantial piece by Ron Rosenbaum in its past May 13 issue. The article gives a good general introduction to an upheaval in textual studies that has been going on for almost forty years, and that has been at white heat for twenty. Because the world of scholarship moves in a kind of slow motion—this remains true even today, odd as that may seem—such belated awareness would not normally be cause for much notice. But at this particular historical moment, when information storage and transmission and methods of knowledge representation are calling for immediate practical attention, Rosenbaum’s piece seems most interesting for what it does *not* talk about. Force of circumstance today calls us to develop scholarly editions in digital forms. The people who have done this work in the past in paper forms—people like Jenkins and Thompson—are involved in serious controversies over how it should be done. The theory and practice of traditional textual scholarship is in a lively, not to say volatile, state of self-reflection. Scholarly editing today cannot be undertaken in *any* medium without a disciplined engagement with editorial theory and method. Scholars who think to use information technology resources, as now we must, therefore face a double difficulty. We must learn to use digital tools whose capacities are still being explored in fundamental ways even by technicians. We must also approach all the traditional questions of scholarly editing as if a transformed world stood all before us, and our choices were fraught with uncertainty. Fortunately, the way will not be a solitary one.

## II . . . . .

To clarify our situation let me rehearse two exemplary recent events. My own work was drawn into the gravity field of both.

Around 1970 various kinds of “social text” theories emerged, pushing literary studies toward a more broadly “cultural” orientation. Interpreters began shifting their focus from “the text” to any kind of social formation in a

broadly conceived discourse field of semiotic works and activities. Because editors and bibliographers oriented their work to physical phenomena—the materials, means, and modes of production—rather than to the readerly text and hermeneutics, this tectonic shift in the larger community of scholars barely registered on the bibliographers’ instruments. A notable exception was D. F. McKenzie, whose 1985 Panizzi Lectures climaxed almost twenty years of work on a social-text approach to bibliography and editing. When they were published in 1986, the lectures brought into focus a central contradiction in literary and cultural studies. Like their interpreter counterparts, textual and bibliographical scholars maintained an essential distinction between empirical/analytic disciplines on one hand, and readerly/interpretive procedures on the other. In his Panizzi Lectures McKenzie rejected this distinction and showed by discursive example why it could not be intellectually maintained.

The distinguished textual scholar T. H. Howard-Hill replied that while views like McKenzie’s were all very well in a theoretical sense, they could not be implemented in a practical way. That is to say, you could not translate such ideas into a scholarly edition. His point was well taken in a paper-based context. Social-text editing proposals commit one to editing books rather than texts—an unfeasible idea in a paper-based view, as Howard-Hill insisted. But digital technology makes such an approach to editing a realizable imagining. One can in fact transform key social and documentary aspects of the book into computable code.

A central purpose of *The Rossetti Archive* project was to prove the correctness of a social-text approach to editing—which is to say, to push traditional scholarly models of editing and textuality beyond the Masoretic wall of the linguistic object we call “the text.” The proof of concept would be the making of the Archive. If our breach of the wall was minimal, as it was, its practical demonstration was significant. We were able to build a machine that organizes for complex study and analysis, for collation and critical comparison, the entire corpus of Rossetti’s documentary materials, textual as well as pictorial. Critical, which is to say computational, attention was kept simultaneously on the physical features and conditions of actual objects—specific documents and pictorial works—as well as on their formal

and conceptual characteristics (genre, metrics, iconography). The Archive's approach to Rossetti's so-called double works is in this respect exemplary. Large and diverse bodies of material that comprise works like "The Blessed Damozel" get synthetically organized: 37 distinct printed texts, some with extensive manuscript additions; 2 manuscripts; 18 pictorial works. These physical objects orbit around the conceptual "thing" we name for convenience "The Blessed Damozel." All the objects relate to that gravity field in different ways, and their differential relations metastasize when subsets of relations among themselves get exposed. At the same time, all of the objects function in an indefinite number of other kinds of relations: to other textual and pictorial works, to institutions of various kinds, to different persons, to varying occasions.

With the Archive one can draw these materials into computable synthetic relations at macro as well as micro levels. In the process the Archive discloses the hypothetical character of its materials and their component parts as well as the relationships one discerns among these things. Though completely physical and measurable (in different ways and scales), neither the objects nor their parts are self-identical, all can be reshaped and transformed in the environment of the Archive.

Don't misunderstand me. Our successes, as I say, have been minimal and some of our greatest hopes for the Archive have not been realized. Nonetheless, the proof of concept was a crucial break with tradition, freeing us to imagine what as yet we don't know: how to build much better and more sophisticated machines of this kind—digital machines that might one day rival that miracle machine; the book *Building the Archive*, for instance, has brought me to realize a possibility for these kinds of instruments that stared us all in the face from the beginning, but that none of us thought to try to exploit. A critical edition can clearly be built in digital form that allows a dynamical tracking and analysis of that recent literary discovery, the "read-erly text." This clearly *also* means that the fundamentally dynamical character of the textual condition can be digitally realized: the dialectic of the field relations between the history of the text's transmission and the history of its reception.

In a late lecture, "What's Past Is Prologue," McKenzie speculated briefly on computerization and textual criticism. His remarks came in the context of two ways that scholars were using digital tools: on one hand for electronic storage of large corpora, on the other for the dynamic modeling of textual materials. McKenzie saw the latter as the more interesting prospect, even if it would "represent a radical departure" from his central "article of bibliographical faith": "the primacy of the physical artifact (and the evidence it bears of its own making)." (There is quintessential McKenzie: entertaining an idea that shook the ground beneath one of his cherished convictions.)

Had he become more involved with the *making* of electronic editions, I believe McKenzie would have realized that, far from departing radically from such primacies, digital tools return us to them in the ways he found most interesting. For "the physical artifact" and "the evidence it bears of its own making" are both social in the sense that such objects, in particular such bibliographical objects, have been made and remade many times in their sociohistorical passages. No book is one *thing*, it is many *things*, fashioned and refashioned repeatedly under different circumstances. Its meaning, as Wittgenstein would say, is in its use. And because all its uses are always invested in real circumstances, the many meanings of any book are socially and physically coded in and by the books themselves. They bear the evidence of the meanings they have helped to make.

One advantage digitization has over paper-based instruments comes not from the computer's modeling powers, but from its greater capacity for simulating phenomena—in this case, bibliographical and sociotextual phenomena. Books are simulation machines as well, of course, with hardcoded machine languages (we call those typography and graphic design) and various softwares (modes of expression—expository, hortatory, imaginative—and genres). The hardware and software of book technology have evolved into a state of sophistication that dwarfs computerization as it currently stands. In time this discrepancy will change, we can be sure. McKenzie probably saw the computer as a modeling machine because of his attachment to "the primacy of the physical object." Computers can be imagined to make models of such primary, self-identical,

objects. But suppose, in our real-life engagements with those physical objects, we experience them as *social objects*, and hence that we see their self-identity as a quantum condition, a function of measurements we choose to make for certain particular purposes. In such a case you will not want to build a model of a made thing, you will try to design a system that can simulate every realizable possibility—the possibilities that are known and recorded as well as those that have yet to be (re)constructed.

McKenzie's central idea, that bibliographical objects are social objects, begs to be realized in digital terms and tools. *The Rossetti Archive* proves that it can be done.

My second example is a cautionary tale that illustrates how that realization can get sidetracked or blocked by a failure to think in clear ways about theory of textuality. The focus of this example is the TEI, the "Text Encoding Initiative," which describes itself as follows:

Initially launched in 1987, the TEI is an international and interdisciplinary standard that helps libraries, museums, publishers, and individual scholars represent all kinds of literary and linguistic texts for online research and teaching, using an encoding scheme that is maximally expressive and minimally obsolescent. (<http://www.tei-c.org/>)

Still an invisible or ghostly presence for many if not most humanities scholars, TEI has become a widely accepted standard for creating electronic texts that require scholarly reliability. It is an inline marking system designed specifically for humanities documents. TEI defines in a precise way an elaborate set of textual information fields so that a computer can search and analyze the texts with respect to those defined fields and extract the marked or "structured" information.

I'm not going to rehearse the problems that have arisen in implementing a TEI approach to machine-readable texts. These were initially aired by the creators of TEI themselves, and subsequent criticisms have confirmed and refined the difficulties. More important to see is the level at which these problems are situated. TEI's greatest legacy is the demonstration it makes of its own inadequacy as a means for computerizing the information content of humanities materials.

TEI understands a text to be "an ordered hierarchy of content objects." This is the same understanding that generated TEI's parent, Standard Generalized Markup Language (or SGML). The view has been criticized, by myself and others, as inadequate for representing the character of poetical and imaginative texts, which mix and overlap various kinds of hierarchical and nonhierarchical features. The criticism, while fairly made, falls far short of exposing the deep inadequacy of an SGML/TEI approach to textuality in the context of digital instruments. It is a criticism, for instance, that can go on to point out—as I have done elsewhere—that if TEI will not do as a markup system for imaginative texts, it will serve nicely for informational texts. That opportunistic position licensed what we did with *The Rossetti Archive*: we used TEI to mark up our informational texts and we developed a special SGML design for all of the Archive's other materials, documentary as well as visual.

But now that we have built the Archive to those design specifications, we can see more clearly the poverty of the result. At such moments Byron's comic wisdom helps you to keep your feet. "In play, there are two pleasures for your choosing, / The one is winning, and the other, losing." The pleasure of losing is what John Unsworth has called, not quite so Byronically perhaps, "The Importance of Failure." The best kinds of defeat come in games that are intense and interesting. Those are the defeats that make you pay, and therefore make you pay attention. Their mythic exemplar is probably the expulsion of Lucifer, the archangel of light and knowledge, from heaven.

As I reflected several years ago on the state of *The Rossetti Archive*, I could see how various practical demands had compromised our initial commitment to the idea of the social text. Most waylaying was our focus on the system's logical design, to the neglect of its interface. We wanted to build a structure that would be, as the digitists say, "bullet-proof" so far as the fast-changing world of hardware and software was concerned. Amazing at it may seem, for six years we built the Archive piece by piece and file by file without ever actually *seeing* anything of the whole except its abstract form: the hypermedia organization of its SGML file structures. For six years the Archive was a soul without a body.

In building digital editions, McKenzie's idea of the social character of physical objects must be held fast. To define a document as a text, as SGML/TEI does, is to follow the rationalist line of textual/bibliographical thinking that McKenzie's work fractured. By contrast, regarding textual documents as physical objects prepares you to develop mechanisms that expose their status as social objects. This is true because physical objects, as McKenzie argued, bear the manifest signs of how and where and by whom they were made. In addition, and reciprocally, physical objects signal their immediate social condition. We can think about ideas and take our solitary way with them. If we fetishize the physical object, we can do the same. But there the move is less easily made because physical objects carry manifest signs of their public and social relations. They have to be handled—that's to say, used and interpreted—with others, in institutional space and in physical ways.

An idea of a Rossetti Archive is not enough, you actually have to make the thing as a physical object. Until you do that you are doomed to what E. P. Thompson once called "the poverty of theory." Postponing—in truth, neglecting—interface design in favor of logical design, the Archive weakened its ability to realize the sociological character and meaning of the physical (social) objects it meant to process. It's easy to see why this result comes about. Logical design is grammatical, interface design is rhetorical. Interface enables and reflects the reader's active presence; it is the environment where readers live and move and have their being in digital simulations.

### III .....

Inadequate as a model for bibliographical things, an SGML/TEI theory of textuality is even less adequate to the processing capabilities of digital instruments. To program a digital information system for hierarchically ordered content objects is to shortchange from the start the simulation capacities of the system. In text-critical terms, it is to design a system that will edit—that will deliver for our use—"texts," not "books," and texts only of a certain, very constrained kind.

These new critical instruments will not suffer for long that kind of dumbing-down. We *can* fashion them to reconstruct an integrated interpretive network of sociological relations for books and other semiotic objects. One type of content object in such a network will be "texts"—that is, linguistic objects formally, as opposed to dialectically, conceived. But we will be calling on these networks to integrate larger masses of different kinds of materials. They will include more even than the bibliographical objects cherished by McKenzie's great Newtonian imagination. I have in mind here what is implicit in the term "interactive," so often—and rightly—applied to digital environments. The critical edition built in digital space interpellates the user as an essential and computable element in the system. The logs that automatically track system usage have scarcely begun to be critically exploited. Skillfully organized, they will develop feedback loops within the network, augmenting the autopoietic mechanisms that are to this point only latent capacities of such systems.

Literary scholars should begin undertaking the serious study of interface design as a necessary modeling preliminary to such work. Interfaces are the mirrors that these systems hold up to their imagined users. Even now we can see—theoretically—that the ideal interface should be as user-specific as possible—more than that, as *use-specific* as possible, for individuals coming to these works may arrive each time with different objects in view. Designing interfaces that are at once stable and flexible, stimulating as well as clear, is one of the two most demanding tasks—in both senses of "demanding"—now facing the scholar who means to work with digital tools. The interfaces should make it clear that when we use a particular machine, we are called to rethink—to change—the territories they initially map for us. We have therefore to see *from* the initial maps that those maps are not precepts but examples of understandings—that they exist to encourage other kinds of mappings and explorations of the material.

The second task insistently before us involves what a traditional humanist would probably see, in no happy Blakean sense, as a marriage of heaven and hell. The work calls together the heavens of literary interpretation and meaning, and the hells of statistics and quantum mechanics.

I can best explain what I mean by reading a passage in a book I recently published. To date,

digital technology has remained instrumental in serving the technical and pre-critical occupations of librarians and archivists and editors. But the general field of humanities education and scholarship will not take the use of digital technology seriously until one demonstrates how its tools improve the ways we explore and explain aesthetic works—until, that is, they expand our interpretational procedures. (*Radiant Textuality*, xii)

I've spent most of my time this afternoon trying to indicate why and how scholarly editions, whether paper or digital, are not the precritical objects that many, probably most, humanities scholars take them to be. That's a theme I've been worrying about and preaching for more than twenty years, which is perhaps a sobering comment on my powers of persuasion. However that may be, the theme returns in this context because most humanists take a similar view of information technology and its relation to the interpretation of cultural works. Computers are the children of a recent science, statistics. Most people who love the humanities hate statistics and so, as with the devil at baptism, we renounce statistics and all its works and all its pomps. At any rate, the statistical devil has been renounced *for us* by our elders.

It's time to stop practicing that secular baptismal rite designed to seal us from what another poet called "too much reality."

Two years ago Johanna Drucker and I began entertaining a way to escape. We would do it with a digital machine we called IVANHOE—named after the once celebrated and massively influential bibliographical romance by Walter Scott, long since, alas, fallen on evil days and evil tongues. We imagined a digitized textual environment—more than that, a discourse field of indefinite extent—that scholars would enter and engage much as people enter and engage with computer games.

You don't perform statistical analyses when you play computer games. You let your servants, the computers, do that for you. And the same is true in IVANHOE, which is a Web-based software application that organizes a collaborative

workspace for research and interpretation of humanities materials, traditional as well as digital. Digitization brings certain advantages to the exploration of such materials. It can simulate a wide variety of informational materials—books, maps, pictures, and so forth—that are the traditional focus of our acts of interpretation. It can access these materials no matter how widely they are dispersed, and it can store, retrieve, reorganize, and transform these massive corpora according to the designs and purposes of specific users.

Deploying synchronous, real-time browser-enabled capabilities in combination with a desktop-based application, IVANHOE thus draws its multiple-user players to seek not so much the "meanings" of these materials as their many possibilities of meaning. IVANHOE multiplies these possibilities in various ways—partly through competition and collaboration between players, partly through the use of masks and roles to constrain the players' interpretive engagements, and partly through immersing the players within a vast field of digitally enhanced and geographically dispersed materials that are specifically organized for further enhancements. We then introduce electronic visualization tools into that field to help us grasp and invent the shapes of thought, both our own and others, as they emerge through our acts of navigating the materials and linking them together in new, imaginative ways.

I've talked often here and abroad about how IVANHOE actually works as an interpretational procedure. Different groups have "played" IVANHOE—if "playing" is the right term—a number of times, including groups of seventh-grade students, college undergraduates, graduate students, and senior humanities scholars that included Johanna Drucker and me. The discourse fields have centered in works like *Wuthering Heights*, *Frankenstein*, *Ivanhoe*, and "The Turn of the Screw." This fall I've brought an elementary model of IVANHOE into a graduate class to test its capacity for enhancing interpretational scholarship in a formal context of graduate research. We're focusing IVANHOE on two distinct scholarly problems: to investigate issues of text and interpretation in Blake's *The Four Zoas*; and to study a set of D. G.

Rossetti's so-called double works in the context of received scholarly ideas about Victorian and modernist aesthetics. This is the first time we've tried to use IVANHOE as a tool for advanced scholarly research.

But the IVANHOE project of exploring new models for interpretational theory and method is a subject in itself and I've already taxed, or perhaps overtaxed, your interest and attention. Let me close this talk with a few remarks about some broad issues of humanities research and scholarship that this digital tool we call IVANHOE—perhaps it is a toy—is raising.

As everyone knows, the scale of information that scholars today are required to negotiate is enormous. Digital instruments have themselves generated—and regenerated—this information in such massive quantities that researchers for some years now have been trying to build quantum computers to handle it. Libraries and museums gather and organize traditional humanities materials in the same way, integrating our received corpora of physical objects like books with our emerging digital corpora. This ever-unfolding informational Archive represents a meta-discourse field, a set of all sets within which we distinguish at our will and choice subset discourse fields that interest us.

When a humanist asks “What *is* this exploding Archive, what is happening here?” part of the answer is that through such an Archive we expose ourselves to ourselves, and our world to itself, in unimaginable depth and detail. But how can we possibly see ourselves or our world—those foundational humanities' goals—in such an information whiteout? Henry Adams's “vehement wish to escape”—a wish he does not indulge, let us remember—turns into Sven Birkerts's advice of refusal.

That very bad advice does little justice to the power and usefulness of the book, which has been our simulation machine of choice for centuries. Now more than ever we want to study the complex mechanisms of book technology in order to design digital environments of comparable sophistication. Think how brilliantly the bibliographical interface organizes our reflective and perceptual experience. It can hold large amounts of different kinds of data and information. At the same time, it sends a clear message that such materials, however rich and strange, are integrated

and negotiable. It facilitates many ways of passaging and repassaging its materials, and of hyperlinking to related materials in and out of books. It leaves us free to understand each in our own ways, and it supplies a bibliographical network ready to receive and feed those diverse readings back into the emergent discourse field.

Compared with that, contemporary digital interface design often seems—often is—less help than hindrance. Bibliography and the sociology of texts are key points of departure for anyone who wants to understand and design digital environments. Reciprocally, digital environments expose the bibliographical discourse field in important new ways. Hypertext, cybertext, ergodic literature: it's true, we have always already been there in our traditional literary forms and functions. But the common reader's view of these comparable technologies is important to remember. People generally think that digital fields are more complex and dynamic than bibliographical ones.

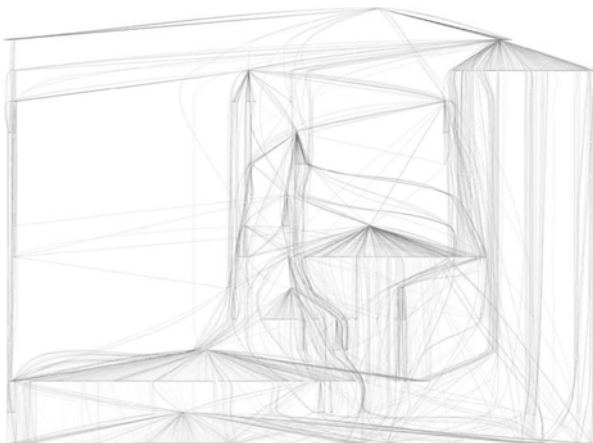
That difference in scale, which is both real and apparent, is important less for its reality than its apparency. I'm not simply being paradoxical in speaking thus. The information whiteout pervading digital space signals poorly designed interface functions. In this context we have much to learn from bibliographical design and the sophisticated information systems to which they are integrated. The codes of simulation operating through printed works are at once robust and amazingly flexible. The passage into digital culture should be made—can only be made, in my opinion—through a reengagement with print culture. It must and will be so because, like Aeneas passing from Troy to Latium, we cannot leave our household gods behind. In this move back to the future we will find ourselves arriving where we started, but now beginning to know that bibliographical place for the first time.

Physicists tell us that a quantum world thunders silently beyond (or below) our human scale of perception. It is a world full of contradictions where everything is as it is perceived, and so everything changes depending on where and how and why you choose to take your observations. In one perspective photons are wave functions, in another they are particles. It is a world of random order and disorder. We were only finally able to establish regular contact with this

world after the invention of statistical mathematics. To the end of his life Einstein disbelieved in the reality of quantum worlds, maintaining they were nothing more than a set of (more or less useful) mathematical functions.

Reality or apparition, a quantum order of bibliographical objects becomes accessible to us through computerization. I am not speaking about the physicochemical makeup of paper objects but of the immense number of dynamic relations and functions that comprise the discourse field of social texts. We touch the hem of this garment whenever we open a Web browser. The field of textual relations accessible through that digital device is statistically significant at a quantum order. People are trying to build quantum computers precisely to improve controlled access to that discourse field.

When such computers are built and made stable enough to be used, history tells us they will have very clumsy interfaces. In the meantime, we have our hands full trying to design interfaces for our current digital tools and systems. We must have them in order to translate the computer's statistical operations into terms that our embodied minds can seize, understand, and put to human uses. The need is especially apparent when the database is a bibliographical discourse field. The interface we have built for *The Rossetti Archive* is dismayingly inadequate to the Archive's dataset of materials. At present the Archive organizes approximately 9,000 distinct files, about half of which are SGML/XML files. When the Archive reaches its scheduled completion date some four years from now, it will have



about twice that many files. Here is a directed graph of a tiny subset of the Archive's current set of analytic relations. We call this "Rossetti Spaghetti."

I show this in order to give you a *graphic* sense of the scale and complexity of this grain of Rossettian sand on the shore of the Internet's opening ocean. One can indeed, even here, see an infinite world dawning in that grain of sand.

Or here is a narrative version of the statistical scale of the Archive. Take those 9,000 files and understand that they are interconnected by a set of some 200,000 hyperlinks. Then add to your equation the fact that every SGML/XML text file is structurally divided into hundreds of types of divisions. Finally, factor in the specific divisionary instances that comprise any particular file, which will range from several hundreds to many thousands. I could ask the server holding the Archive to make the actual counts in each case, but I think you can see the staggering number of possible relationships that the Archive puts into computational play.

Let me close with what is for me—a fetishist of imaginative writing, especially poetry—the most important moral of the whole story: that poems and other imaginative kinds of social texts are quantum fields. Although we have said for a long time that their meanings are inexhaustible, pursuing a sociology of textuality in a digital frame of reference helps us to specify more clearly why and how this is the case. I do not offer the quantum poem as a useful metaphor but as a fact about the facts comprising poetic discourse fields—a computable fact. The implications of that view of social textuality for humanities studies seem to me considerable. IVANHOE is an early effort to work out those implications for a program of what I. A. Richards once called "Practical Criticism." Johanna Drucker and I call it "Pataphysical Criticism." Like Byron, Alfred Jarry's ludic intelligence is (so to say) no joking matter. From Ubu and Dr. Faustroll emerges an algorithmic form of scholarly method that should be seriously entertained (so to say). It is, I believe, the only method adequate to the textual condition we now see clearly unfolding before us.



# Promises and Perils of Digital History

Roy Rosenzweig

*This paper was originally presented as the Richard W. Lyman lecture at the National Humanities Center on December 4, 2003. It was later revised and published as the introduction to Digital History: A Guide to Gathering, Preserving, and Presenting the Past on the Web (Philadelphia: University of Pennsylvania Press, 2006). That book was coauthored with Daniel J. Cohen, and I want to express my deep thanks for his collaboration and contributions, which greatly improved the original talk. This revised version of the introduction is reprinted with the permission of University of Pennsylvania Press.*



Step back in time and open the pages of the inaugural issue of *Wired* magazine from the spring of 1993 and prophecies of an optimistic digital future call out to you. Management consultant Lewis J. Perleman confidently proclaims an “inevitable” “hyperlearning revolution” that will displace the thousand-year-old “technology” of the classroom, which has “as much utility in today’s modern economy of advanced information technology as the Conestoga wagon or the blacksmith shop.” John Browning, a friend of the magazine’s founders and later the Executive Editor of *Wired UK*, rhapsodizes about how “books once hoarded in subterranean stacks will be scanned into computers and made available to anyone, anywhere, almost instantly, over high-speed networks.” Not to be outdone by his authors, *Wired* publisher Louis Rossetto links the digital revolution to “social changes so profound that their only parallel is probably the discovery of fire.”<sup>1</sup>

While the *Wired* prophets could not contain their enthusiasm, the techno-skeptics fretted about a very different future. Debating *Wired* Executive Editor Kevin Kelly in the May 1994 issue of *Harper’s*, literary critic Sven Birkerts implored readers to “refuse” the lure of “the electronic hive.” The new media, he warned, pose a dire threat to the search for “wisdom” and “depth”—“the struggle for which has for millennia been central to the very idea of culture.”<sup>2</sup>

Some historians—on both the right and the left—also saw deep trouble ahead. In November 1996, the conservative Gertrude Himmelfarb offered what she called a

“neo-Luddite” dissent about “the new technology’s impact on learning and scholarship.” “Like postmodernism,” she complained, “the Internet does not distinguish between the true and the false, the important and the trivial, the enduring and the ephemeral.... Every source appearing on the screen has the same weight and credibility as every other; no authority is ‘privileged’ over any other.” A year later, the Marxist historian of technology David Noble found himself standing beside Himmelfarb in the neo-Luddite crowd, although not surprisingly he spotted the cyberthreat coming from a different direction. “A dismal new era of higher education has dawned,” he wrote in a paper called “Digital Diploma Mills: The Automation of Higher Education.” “In future years we will look upon the wired remains of our once great democratic higher education system and wonder how we let it happen.”<sup>3</sup>

More than a decade into the promised “digital revolution,” the cyber-enthusiasts and the techno-skeptics have both turned out to be poor prophets of the future. Universities and libraries still stand. Culture has not crumbled. Paradise has not arrived. But to decide that neither utopia nor dystopia beckons should not lead to the comfortable conclusion that nothing has changed or will change. Driven by the rapid emergence and dissemination of computers, global computer networks, and new digital media, change—though not revolution—surrounds us. Our daily habits of finding the news and weather, buying books, and communicating with colleagues and loved ones have permanently changed.

Even the ancient discipline of history has begun to metamorphose. In the past two decades new media and new technologies have challenged historians to rethink the ways that they research, write, present, and teach about the past. Almost every historian regards a computer as basic equipment; colleagues view those who write their books and articles without the assistance of word-processing software as objects of curiosity. History teachers labor over their PowerPoint slides as do sixth graders preparing for History Day. Email and instant messaging has broadened circles of communication and debate among dispersed historical practitioners, scholars as well as amateur enthusiasts.

Nowhere are the signs of change for historians more evident than on the World Wide Web. *Yahoo’s* web directory currently lists more than 32,000 history websites. Even this vast catalog greatly underestimates the pervasiveness of the past online, not including, for example, the tens of thousands of online syllabi for history courses. In the past decade, historians with interests ranging from ancient Mesopotamia to the post-Cold War world have enthusiastically embraced the web. Virtually every scholarly journal duplicates its content online (though not always openly), and almost every history course has its syllabus posted on the web. Virtually every historical archive, historical museum, historical society, historical house, and historic site—even the very smallest—has its own website. So does just about every re-enactment group, genealogical society, and body of historical enthusiasts.

My own position as founder and director of the Center for History and New Media and recipient of the Lyman Award implicitly puts me on the other side of the fence from the neo-Luddite historians like Noble and Himmelfarb. I obviously believe that we gain something from “doing digital history,” making use of the new computer-based technologies. Yet while skeptical of the arguments of the techno-skeptics, I am not entirely enthusiastic about the views of the cyber-enthusiasts either. Rather, I believe that we need to critically and soberly assess where computers, networks, and digital media are and aren’t useful for historians—a category that I define broadly to include amateur enthusiasts, research scholars, museum curators, documentary filmmakers, historical society administrators, classroom teachers, and history students at all levels. In what ways can digital media and digital networks allow us to do our work as historians better?

This talk briefly sketches seven qualities of digital media and networks that potentially allow us to do things better: capacity, accessibility, flexibility, diversity, manipulability, interactivity, and hypertextuality (or nonlinearity). I also talk about five dangers or hazards on the information superhighway: quality, durability, readability, passivity, and monopoly. This scorecard of possibilities and problems seems, on balance, to suggest a digital future worth pursuing. I thus align myself with neither the wild-eyed optimists

nor the gloomy pessimists but rather with the camp known as “techno-realists” who seek, in the words of computer scientist and social theorist Phil Agre, to analyze “case-by-case the interactions between technology and institutions through which the action really unfolds.”<sup>4</sup> Doing digital history well entails being aware of the technology’s advantages and disadvantages, and how to maximize the former while minimizing the latter.

The first advantage of digital media for historians is storage *capacity*—digital media can condense unparalleled amounts of data into small spaces. A 100-gigabyte portable hard drive that sells for \$99 and weighs 7 ounces can hold a 100,000-volume library. Since historians love data and archival sources, they have great interest in this ability to condense large amounts of data into tiny amounts of space. Historians who would like to make large quantities of primary sources available over the web quickly learn that storage space is perhaps the smallest expense they face.

The most profound effect, however, may be on tomorrow’s historians. The rapidly dropping price of data storage has led computer scientists like Michael Lesk (a cyber-enthusiast to be sure) to claim that in the future, “there will be enough disk space and tape storage in the world to store everything people write, say, perform or photograph.” In other words, why delete anything from the current historical record since it costs so little to save it? What might it mean to write history when no historical evidence has disappeared?<sup>5</sup>

The vast storage capacity of digital media would be of much less interest without a second and even more important advantage—its *accessibility*. This quality derives both from the ability to condense the bits and bytes encoded in digital media into small spaces but even more from the emergence of ubiquitous computer networks that can almost instantly send those bits around the world. Historians have multiple audiences; digital networks mean that we can reach those audiences—students, other scholars and teachers, the general public—much more easily and cheaply than ever before. The distribution of history projects electronically approaches what the economists call “zero marginal cost”; once the initial expenses are met, reaching an additional person costs almost nothing (unlike, say, a print book

where costs decline after the initial investment but still remain substantial). Our web server at the Center for History and New Media (CHNM) gets about three-quarters of a million hits a day, but on September 11, 2002 (when people looking to commemorate the attacks of the previous year descended in droves on the September 11 Digital Archive that the Center for History and New Media organized in collaboration with the American Social History Project), we handled 8 million hits—a tenfold increase with no additional costs.<sup>6</sup>

Online accessibility means, moreover, that the documentary record of the past is open to people who rarely had entrée before. The analog Library of Congress has never welcomed high school students—its reading rooms, no less its special collections, routinely turn them away. Now the library’s American Memory website allows high school students to enter the virtual archive on the same terms of access as the most senior historian or member of Congress. To those who previously had no easy access, online archives open locked doors. Non-academic users of the University of North Carolina’s archival website, *Documenting the American South*, reports university librarian Joe Hewitt, speak eloquently of how they “felt privileged to have access to these primary sources as if they had entered an inner sanctum where they did not fully belong.”<sup>7</sup> But even for well-credentialed historians such online archives put millions of historical documents at hand twenty-four hours a day and without the cost of a plane ticket to Washington, D.C. or Chapel Hill, North Carolina—and without the delay. The instantaneous access to primary and secondary sources—the ability to very quickly make and test out intellectual connections—will likely alter historical research and writing in ways that we haven’t yet imagined.

The accessibility and publicness of the web have consequences for history projects much less extensive than those mounted by the Library of Congress or major university libraries. High school teachers can devise community history projects in which students present the results of their research to a public audience of local residents. Historical societies of small and declining towns on the Great Plains can keep in touch with—and gather historical information from—former residents.<sup>8</sup> A genealogical

web page can bring together the descendants of a family who started out in County Cork, Ireland but later scattered to London, Toronto, San Francisco, Cape Town, and Melbourne. The Internet allows historians to speak to vastly more people in widely dispersed places without really spending more money—an extraordinary development.

The past that is suddenly more accessible is also much richer because of a third characteristic of digital media—what we might call its *flexibility*. Because digital media are expressed in a basic language of 1s and 0s, they can take multiple forms, and that means we can arrange those bits into text, images, sounds, and moving pictures. Thus, we can more easily preserve, study, and present the past in the multiple media that expressed and recorded it. The online digital archives can contain images, sounds, and moving pictures as well as text. And you can present the past in multiple media that *combine* sounds, images, and moving pictures with words.

But the flexibility of digital data lies not just in its ability to encompass both text and audio. It also resides in the ability of the same data to take multiple forms automatically. Although language translation software is still primitive, we are moving toward a time when words in one tongue can be automatically translated into another—perhaps not perfectly but effectively enough. More generally, digital information organized into databases or marked up in structured languages like XML can be instantly reordered or combined into new forms. Acting on the pieces in a database or XML document, small but powerful computer programs can pull together disparate materials in a way that compares, contrasts, and enhances them. For example, a scholar of ancient Greece simultaneously can see an image of a vase, commentaries from several other historians about that vase, and suggestions of similar artifacts from a database. As new media theorist Lev Manovich points out, the “numerical coding of media” and the “modular structure of a data object” mean “a new media object is not something fixed once and for all, but something that can exist in different, potentially infinite versions.” Thus, Manovich sees the database—with its infinitely rearrangeable data—as one of the fundamental forms found in new media.<sup>9</sup>

Flexibility transforms the experience of consuming history, but digital media—because of its openness and *diversity*—also alters the conditions and circumstances of producing history. The computer networks that have come together in the World Wide Web are not only more open to a global audience of history readers than any other previous medium, they are also more open to history authors. A 2004 study found that almost half of Internet users in the United States have created online content by building websites, creating blogs, and posting and sharing files. An astonishing 13 percent maintain their own websites, and one recent census counts 66 million blogs.<sup>10</sup> No publishing medium has ever had such a low barrier to entry. At virtually no cost millions have access to their own printing press. Already, the number of authors of history web pages is likely greater than the number of authors of history books. But the even more dramatic contrast is in the social composition of the two sets of authors—web history authors are significantly more diverse and significantly less likely to have formal credentials. Their strong presence online unsettles existing hierarchies, thus producing Himmelfarb’s jeremiad and the laments of other techno-skeptics.

The web, as a result, has given a much louder and more public voice to amateur historians. If you put “Abraham Lincoln” in Google, one of the first sites listed is the *Abraham Lincoln Research Site*, which features the writing of Roger Norton, who says of himself “I am not an author or an historian; rather I am a former American history teacher who enjoys researching Abraham Lincoln’s life and accomplishments.”<sup>11</sup> Through Google’s eye, which is how an increasing number of people view the web, Roger Norton is a more influential Lincoln historian than the Pulitzer-prize winning Harvard professor David Donald.

For the most part, these first four qualities of digital media provide what we might call quantitative advantages—we can do more, reach more people, store more data; give readers more varied sources; we can get more of them into classrooms; give students access to more materials, hear from more perspectives. But does digital history do anything differently? Literary critic Janet Murray raises this issue in *Hamlet on the Holodeck*, her book on the future of narrative in cyberspace. There, she distinguishes

between “additive” and “expressive” features of new media. She makes the useful analogy to early films, which were initially called “photoplays,” and thus thought of as “a merely additive art form (photography plus theatre).” Only when filmmakers learned to use montage, close-ups, zooms and the like as part of storytelling did photoplays give way to the new expressive form of movies.<sup>12</sup>

To consider these “expressive” qualities we need to think, for example, about *manipulability* of digital media—the possibility of manipulating historical data with electronic tools as a way of finding things that were not previously evident. At the moment, the most powerful of those tools for historians is the simplest—the ability to search through vast quantities of text for particular strings of words. The word search capabilities of JSTOR, the online database of 460 scholarly periodicals, makes possible a kind of intellectual history that cannot be done as readily in print sources. Say you want to trace the changing reputation of Richard Hofstadter in the historical profession; the 667 articles in JSTOR that mention Hofstadter provide an invaluable starting point. Historians of language are already having a field day playing with such massive databases. The librarian and lexicographer Fred Shapiro, for example, has uncovered uses of such phrases as “double standard” (1912) and “Native American” (for American Indian, 1931) that predate citations in the *Oxford English Dictionary* by decades. Similarly, CHNM’s Syllabus Finder makes it possible to discover—by searching through thousands of online history syllabi—patterns in history teaching (the popularity of different courses, texts, or types of assignments) that were previously invisible.<sup>13</sup>

But text searching is only one very simple technique, albeit a powerful one when leveraged through Boolean searches and the use of advanced pattern-matching methods that computer scientists call “regular expressions.” Even more tantalizing are the prospects of being able to search automatically through vast quantities of images, sounds, and moving pictures. And, at some point, we may be able to dynamically map (temporally and geographically) historical events drawn from tens of thousands of historical sources. Or we may be able to see new things in historical images through digital close-ups or manipulation. Jerome

McGann, for example, talks about using software tools to “deform” images and see in them elements previously missed.<sup>14</sup>

Digital media also differs from many other older media in its *interactivity*—a product of the web being, unlike broadcast television, a two-way medium, in which every point of consumption is a point of production. This interactivity enables multiple forms of historical dialogue—among professionals, between professionals and nonprofessionals, between teachers and students, among students, among people reminiscing about the past—that were possible before but which are not only simpler but potentially richer and more intensive in the digital medium. Many history websites offer opportunities for dialogue and feedback. The level of response has varied widely but the experience so far suggests how we might transform historical practice—the web becomes a place for new forms of collaboration, new modes of debate, and new modes of collecting evidence about the past. At least potentially, digital media transforms the traditional, one-way reader/writer, producer/consumer relationship. Public historians, in particular, have long sought for ways to “share authority” with their audiences; the web offers an ideal medium for that sharing and collaboration. Probably the most powerful example of this participatory history writing that “shares authority” is the open source encyclopedia Wikipedia, which has been written largely by nonprofessionals and has become the most important online historical reference.<sup>15</sup>

Finally, we note the *hypertextuality* or *nonlinearity* of digital media—the ease of moving through narratives or data in undirected and multiple ways. Hypertext, as is well known, is a constitutional principle of the World Wide Web; its original designer, Tim Berners-Lee, called its most basic protocol the “Hypertext Transfer Protocol”—the “http” that begins every web address. For postmodernists, hypertextuality fractures and decenters traditional master narratives in beneficial ways. “Hypertext,” writes literary critic George Landow, “emphasizes that the marginal has as much to offer as the central by refusing to grant centrality to anything...for more than the time a gaze rests upon it. In hypertext, centrality, like beauty and relevance, resides in the eye of the beholder.” For Landow, hypertext reconfigures texts,

authors, writing, and narrative. In this fundamental “paradigm shift” (what he calls “a revolution in human thought”), conceptual systems “founded upon ideas of center, margin, hierarchy, and linearity” are overturned by “ones of multilinearity, nodes, links, and networks.”<sup>16</sup>

To talk about revolutions in human thought starts to make us sound like one of the cyber-enthusiasts with whom I began. Are we, in fact, on the verge of a new, richer, and rewarding era of cyber-historical work—a digital history revolution? While I would not disavow the profound advantages and features of digital history, I would quickly offer some caveats. Some equally profound barriers and difficulties keep all of us from reaching this rosy digital future. Moreover, some of the positive goods that online history is bringing to our desktops are accompanied by serious hazards and dangers—many of them are, in turn, the flip side of advantages I discussed earlier.

For example, the problems of *quality* and authenticity emerge, in part, out of the vast capacity of digital media. Often cyber-skeptics summarize this view in the simple phrase “it’s mostly junk.” “Internet search engines,” writes Gertrude Himmelfarb, “will produce a comic strip or advertising slogan as readily as a quotation from the Bible or Shakespeare.” Historian James William Brodman similarly worries that students will unfailingly grab the comic strip rather than Shakespeare: “Much of the material that students...unearth in cyberspace is of uneven character—juvenile, inaccurate, or sometimes simply wrong.”<sup>17</sup>

And to be sure, we can find plenty of inaccurate history on the web. Take a look at the web pages of Citizens for a Sound Economy and the Federal Reserve Bank of Dallas and read a letter allegedly from Martin Van Buren to Andrew Jackson calling for government intervention to stop the threat to the railroads posed by the Erie Canal. A careful assessment of internal evidence (an important historical skill in all ages) readily betrays the twentieth-century origins of this “nineteenth-century” letter. But the forgery predates the web, and the web also offers crucial evidence about the origins of the counterfeit. Moreover, in general, the web is more likely to be right than wrong. A quick check of Google finds 613 web pages discussing the “Gettysberg Address” but 86,100 that correctly spell the locale for

Lincoln’s speech as “Gettysburg.” If the existence of misinformation on the web is no more of a problem than its existence in the rest of American society, the web does actually pose some thornier problems of authenticity and authority. One is that both forgery and the movement of forgery into the “information stream” is considerably easier in the digital and networked world.<sup>18</sup>

Consider, for example, the famous “photograph” of Lee Harvey Oswald and Jack Ruby playing rock music together in a Dallas basement. Such fake photographs have a long history; Stalin’s photo retouchers, for example, spent considerable time airbrushing Trotsky out of the historical record. But the transformation of the original Bob Jackson photo of Ruby shooting Oswald into “In-A-Gadda-Da-Oswald” did not require a skilled craftsman. George Mahlberg created it with Photoshop in forty minutes and it quickly spread across the World Wide Web, popping up in multiple contexts that erase the credit of the “original” counterfeiter.<sup>19</sup>

Himmelfarb implies a related problem in her horror that a comic strip could have the same authority as the Bible. In this new space, will traditional repositories of authority retain their stature and influence? In the heterogeneous space of the web, will the History Channel serve as a more influential authority than the History Cooperative, the online publisher of the *American Historical Review* and the *Journal of American History*? Anyone can gain admittance to the History Channel site, but the History Cooperative site is only open to journal subscribers.

Is there some way to police the boundaries of historical quality and authenticity on the web? Could we stop a thousand historical flowers—amateur, professional, commercial, crackpot—from blooming on the web? Would we want to? Of course, issues of quality, authenticity, and authority predate the Internet. But digital media undercut an existing structure of trust and authority and we, as historians and citizens, have yet to establish a new structure of historical legitimation and authority. When you move your history online, you are entering a less structured and controlled environment than the history monograph, the scholarly journal, the history museum, or the history classroom. That can have both positive and unsettling implications.

One vision of the digital future involves the preservation of everything—the dream of the complete historical record. The current reality, however, is closer to the reverse of that—we are rapidly losing the digital present that is being created because no one has worked out a means of preserving it. The flip side of the flexibility of digital data is its seeming lack of *durability*—a second hazard on the road to digital history nirvana. The digital record of the federal government is being lost on a daily basis. Although most government agencies started using email and word processing software in the mid-1980s, the National Archives still does not require that digital records be stored in their original form. Nor are there any archiving guidelines for the 26 million U.S. government web pages.<sup>20</sup> Again, historical and archival preservation are hardly new problems but the digital era has forced us to reconsider fundamental questions about what should be preserved and who should preserve it.

Prophets of hypertext have repeatedly promised a new, richer reading experience, but critics have instead seen the digital environment as the death of reading, as we know it. Sven Birkerts has expressed the most profound sense of loss in *Gutenberg Elegies: The Fate of Reading in an Electronic Age*. The more prosaic (and the most common) complaint centers on the difficulty of reading a screen. But reading on screen may ultimately find a technological solution as high-resolution, high-contrast screens become cheaper to produce.<sup>21</sup>

The more profound problem of *readability* is figuring out what it means to be an author in this environment. Typically, such experiments place large demands on the reader—they are, in Espen Aarseth's phrase "ergodic" literature, in which "nontrivial effort is required to allow the reader to traverse the text." Historian Philip J. Ethington's online article on Los Angeles—the *American Historical Review's* first all-electronic work—asks you to make your way through a relatively dense argument for a spatial theory of historical certainty as well as a vast set of videos, panoramas, maps, and essays on everything from photography to urban epistemology.<sup>22</sup>

Hypertext scholarship like this disrupts the conventions of the printed scholarly article. Yet while such conventions

can be deadening, they can also make printed articles easy to read, at least by those who know the "codes." Most academics can rapidly find the thesis in the first few pages, the conclusions on the last two pages, and a sense of the sources used through a quick scan of the footnotes. Such strategies are worthless in confronting hypertext essays. Not only is the thesis often hard to find quickly, but it is not always clear that there *is* a thesis. Where is the beginning? The end? Reader expectations about the investment of time required to master an essay are entirely disrupted. In effect, those works undercut the unwritten social contract that exists between readers and writers of scholarly essays—a social contract in which the author agrees to follow conventions of argumentation, organization, and documentation, and the reader agrees to devote a certain amount of time to give the article a fair reading.<sup>23</sup>

Digital enthusiasts assume that the online environment is intrinsically more "interactive" than one-way, passive media like television. But digital technology could, in fact, foster a new couch potato-like *passivity*. Such preferences make efforts at interactive history projects particularly quixotic when they also must confront the fact that computers are good at yes and no, right and wrong, and historians prefer words like "maybe," "perhaps," and "it is more complicated than that." Thus the most common form of historical interactivity on the web is the multiple-choice test. But the high-budget version is little better. Take, for example, the History Channel's website *Modern Marvels Boy's Toys*, which is a combination of watching the cable channel and playing a video game. The true interactivity here comes when you click on the "shop" button. As legal scholar Lawrence Lessig has written pessimistically: "There are two futures in front of us, the one we are taking and the one we could have. The one we are taking is easy to describe. Take the Net, mix it with the fanciest TV, add a simple way to buy things, and that's pretty much it." At the same time, some wonder whether we really want to foster "interactivity" at all, arguing that it fails to provide the critical experience of understanding, of getting inside the thoughts and experiences of others. The literary critic Harold Bloom, for example, argues that whereas linear fiction allows us to experience more by granting us access to the lives and thoughts

of those different from ourselves, interactivity only permits us to experience more of ourselves.<sup>24</sup>

A more serious threat in digital media, which runs counter to its great virtues of accessibility, publicity, and diversity, is the real potential for *inaccessibility* and *monopoly*. The best-known danger—the digital divide in computer ownership and Internet use between rich and poor, black and white—has diminished somewhat but persists despite politically motivated claims to the contrary. And on a global basis, the divide is wide indeed; two-thirds of the people in the world have no access to telephones, let alone the Net. Moreover, even as more and more people acquire computers and Internet connections, they do not simultaneously acquire the skills for finding and making effective use of this new, free global library.<sup>25</sup> Another concern stems more from the production than the consumption side. Will amateur and academic historians be able to compete with well-funded commercial operators—like the History Channel—for attention on the Net?

In any event, the most important commercial purveyors of the past are not, at the moment, the History Channel or TheHistoryNet but global multibillion-dollar information conglomerates like Reed Elsevier and the Thomson Corporation, which charge libraries high prices for the vast digital databases of journals, magazines, newspapers, books and historical documents that they control.<sup>26</sup> Dividing cyberspace into a series of gated communities controlled by information conglomerates means that the dream of a globally interconnected scholarship is just that—a dream. The balkanization of the web into privately owned digital storehouses has been made worse by the scandalous Sonny Bono Copyright Extension Act of 1998, which extended existing copyrights by another twenty years (in part due to the aggressive lobbying of the Disney Corporation, whose Mickey Mouse was scurrying toward the public domain). Will “authority” and “authenticity” reside with the corporate purveyors of the past? Will the diverse, eclectic, and largely free public history web survive the onslaught of these mega operations? Will access to the best historical resources be open or closed?

Such questions and concerns should not lead us to throw up our hands in despair. Rather they should prod us

to sit down in front of our computers and get to work. Historians need to confront these issues of authority, durability, readability, passivity, and inaccessibility rather than leave them to the technologists, legislators, and media companies, or even just to our colleagues in libraries and archives. We should put our energies into maintaining and enlarging the astonishingly rich public historical web that has emerged in the past decade. For some, that might involve joining “the international effort to make research articles in all academic fields freely available on the Internet,” as embodied, for example, in the Budapest Open Access Initiative.<sup>27</sup> For others, that should mean joining in eclectic but widespread grass-roots efforts to put the past online—whether that involves posting a few documents online for your students or raising funds for more ambitious projects to create free public archives. Just as “open source” code has been the banner of academic computer scientists, “open sources” should be the slogan of academic and popular historians. Academics and enthusiasts created the web; we should not quickly or quietly cede it to giant corporations. The most important weapon for building the digital future we want is to take an active hand in creating digital history in the present.

## Notes

<sup>1</sup> Lewis J. Perelman, “School’s Out: The Hyperlearning Revolution Will Replace Public Education”; John Browning, “Libraries Without Walls for Books Without Pages”; and Louis Rosetto, “Why Wired?” all in *Wired Magazine*, March–April 1993, parts of which are available at <http://www.wired.com/wired/archive/1.01/>.

<sup>2</sup> Sven Birkerts in “The Electronic Hive: Two Views,” *Harper’s Magazine*, May 1994, 17–21, 24–25. See also Birkerts, *The Gutenberg Elegies: The Fate of Reading in an Electronic Age* (Boston: Faber & Faber, 1994).

<sup>3</sup> Gertrude Himmelfarb, “A Neo-Luddite Reflects on the Internet,” *Chronicle of Higher Education*, November 1, 1996, A56; Noble’s essay was later reprinted in David Noble, *Digital Diploma Mills: The Automation of Higher Education* (New York: Monthly Review Press, 2001).

<sup>4</sup> Phil Agre, “[RRE] Notes and Recommendations,” email to “Red Rock Eater News Service,” [rre@lists.gseis.ucla.edu](mailto:rre@lists.gseis.ucla.edu), 8 August 2000, <http://commons.somewhere.com/rre/2000/RRE.notes.and.recommenda11.html>.

<sup>5</sup> Michael Lesk, “How Much Information Is There in the World?” *Michael Lesk*, <http://www.lesk.com/mlesk/ksg97/ksg.html>. See also Roy Rosenzweig, “Scarcity or Abundance? Preserving the Past in a Digital Era,” *American Historical Review* 108, 3 (June 2003): 735–762, <http://chnm.gmu.edu/assets/historyessays/scarcity.html>.

<sup>6</sup> CHNM and ASHP, *The September 11 Digital Archive*, <http://911digitalarchive.org>. Of course, a long-term increase in traffic of that scale will increase costs (for bandwidth and for support, for example) but not to the same degree as, say, printing sixteen-times as many copies of a book.

<sup>7</sup> Library of Congress, "American Memory: Historical Collections for the National Digital Library," <http://memory.loc.gov/>; Academic Affairs Library, University of North Carolina at Chapel Hill, "Documenting the American South," <http://docsouth.unc.edu/>; Joe A. Hewitt, "Remarks," Doc South One-thousandth Title Symposium, Chapel Hill, North Carolina, March 1, 2002, <http://docsouth.unc.edu/jahewitt.html>.

<sup>8</sup> See, for example, Kevin Roe's website "Brainerd, Kansas: Time, Place and Memory on the Prairie Plains," <http://www.rootinaround.com/brainerd/>.

<sup>9</sup> Lev Manovich, *The Language of New Media* (Cambridge, Mass.: MIT Press, 2001), 36, 214. As Jim Sparrow has pointed out to us, this, in effect, reverses Walter Benjamin's famous argument in "The Work of Art in the Age of Mechanical Reproduction" that an "authentic" original is essential to our valuation of art as a "priceless" work of human creativity. For Benjamin, "mechanical reproduction" erased value but here it unleashes the collective process of interpretation, debate, and memory from which genuine historical value flows. In other words, access harnesses and democratizes the collective work that undergirds the production of historical knowledge.

<sup>10</sup> Amanda Lenhart, John Horrigan, and Deborah Fallows, *Content Creation Online* (Washington, DC: Pew Internet and American Life Project, 2004), [http://www.pewinternet.org/PPF/r/113/report\\_display.asp](http://www.pewinternet.org/PPF/r/113/report_display.asp); *Technorati* currently tracks 66 million blogs, <http://technorati.com/about/>.

<sup>11</sup> Roger Norton, *Abraham Lincoln Research Site*, <http://members.aol.com/RVSNorton/Lincoln2.html>. Norton was first between June 2003 and July 2004 but had slipped to number three by August.

<sup>12</sup> Janet H. Murray, *Hamlet on the Holodeck: The Future of Narrative in Cyberspace* (New York: Free Press, 1997), 66.

<sup>13</sup> Ethan Bronner, "You Can Look It Up, Hopefully," *New York Times*, January 10, 1999; Daniel J. Cohen, "By the Book: Measuring the Importance of Textbooks to the U.S. Survey Courses That Use Them," *Journal of American History*, forthcoming March 2005.

<sup>14</sup> Jerome McGann, *Radiant Textuality: Literature After the World Wide Web* (New York: Palgrave, 2001), 84-85. For a brief overview of the state of content-based image retrieval, see Humanities Advanced Technology and Information Institute and National Initiative for a Networked Cultural Heritage, *The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials*, 45, <http://www.nyu.edu/its/humanities/ninchguide/>.

<sup>15</sup> Michael Frisch, *A Shared Authority: Essays on the Craft and Meaning of Oral and Public History* (Albany: State University of New York Press, 1990). Roy Rosenzweig and David Thelen, *Presence of the Past: Popular Uses of History in American Life* (New York: Columbia University Press, 1998). On Wikipedia, see Roy Rosenzweig, "Can History Be Open Source? Wikipedia and the Future of the Past," *Journal of American History* 93, no. 1 (June 2006), <http://mutex.gmu.edu:2069/journals/jah/93.1/rosenzweig.html>.

<sup>16</sup> George P. Landow, *Hypertext 2.0* (Baltimore: Johns Hopkins University Press, 1997), 89, 2.

<sup>17</sup> Himmelfarb, "A Neo-Luddite Reflects on the Internet"; James William Brodman, "E-Publishing: Prospects, Promises, and Pitfalls," *AHA Perspectives* (February 2000), <http://www.theaha.org/perspectives/issues/2000/0002/0002vie1.cfm>.

<sup>18</sup> "Tech Bytes - Tid Bits in Tech News: Endangering Life and Limb... At Breakneck Speed," *Citizens for a Sound Economy*, [http://www.cse.org/informed/issues\\_template.php?issue\\_id=348](http://www.cse.org/informed/issues_template.php?issue_id=348); Bob McTeer, "The Great Trade Debates and What's at Stake," Remarks delivered at the World Affairs Council and Texas International Trade Alliance, Houston, Texas, 10 October 2000, <http://www.dallasfed.org/mcteer/speeches/2000/ms001010.html>; David Bearman and Jennifer Trant, "Authenticity of Digital Resources: Towards a Statement of Requirements in the Research Process," *D-Lib Magazine* 4, no. 6 (June 1998), <http://www.dlib.org/dlib/june98/o6bearman.html>. For a more systematic discussion of how automated methods can find information of high quality on the web, see Daniel J. Cohen and Roy Rosenzweig, "Web of Lies? Historical Knowledge on the Internet," *First Monday* 10 (December 2005), [http://firstmonday.org/issues/issue10\\_12/cohen/index.html](http://firstmonday.org/issues/issue10_12/cohen/index.html).

<sup>19</sup> One website with this photo is ArteMedia, "In-A-Gadda-Da-Oswald," <http://www.artemedia.com/ephemera/oswald.html>.

<sup>20</sup> Rosenzweig, "Scarcity or Abundance?"; Jeffrey Benner, "Is U.S. History Becoming History?" *Wired News* (9 April 2001), <http://www.wired.com/news/politics/0,1283,42725,00.html>.

<sup>21</sup> Birkerts, *Gutenberg Elegies*; Roy Rosenzweig, "Crashing the System: Hypertext and American Studies Scholarship," *American Quarterly* 51 (June 1999): 237-46. Business historian Austin Kerr wrote in an online response to the online AQ essays: "a few moments was enough...I certainly did not feel comfortable trying to read this one example." Sony's recently (2004) released Librie, which offers resolution of 600x800 dots at 170dpi, provides a glimpse of things to come. J. Mark Lytle, "Library Without Books," *Guardian* (22 April 2004).

<sup>22</sup> Espen J. Aarseth, *Cybertext: Perspectives on Ergodic Literature* (Baltimore: Johns Hopkins University Press, 1997); Philip J. Ethington, "Los Angeles and the Problem of Urban Historical Knowledge," *American Historical Review* 105 (December 2000), <http://cwis.usc.edu/dept/LAS/history/historylab/LAPUHK/index.html>.

<sup>23</sup> See Rosenzweig, "Crashing the System."

<sup>24</sup> <http://www.historychannel.com/boystoys/>; Lawrence Lessig, *The Future of Ideas: The Fate of the Commons in a Connected World* (New York: Random House, 2001), 7; Harold Bloom, *How to Read and Why* (New York: Scribners, 2001).

<sup>25</sup> On the digital divide see Mark N. Cooper, *Does the Digital Divide Still Exist? Bush Administration Shrugs, But Evidence Says Yes* (Washington, DC: Consumer Federation of America, 2002), <http://www.consumerfed.org/DigitalDivideReport20020530.pdf>; Jeffrey Benner, "Bush Plan 'Digital Distortion,'" *Wired News*, 7 February 2002, <http://www.wired.com/news/politics/0,1283,50279,00.html>; Eszter Hargittai, "Second-Level Digital Divide: Differences in People's Online Skills," *First Monday* 7, no. 4 (April 2002), [http://firstmonday.org/issues/issue7\\_4/hargittai/index.html](http://firstmonday.org/issues/issue7_4/hargittai/index.html).

<sup>26</sup> Roy Rosenzweig, "The Road to Xanadu: Public and Private Pathways on the History Web," *Journal of American History* 88, no. 2 (2001): 548-579, <http://chnm.gmu.edu/assets/historyessays/roadtoxanadu.html>. For example, the list price for Thomson Corporation's eighteenth-century digital collection begins at \$500,000: Kinley Levack, "Digital ECCOs of the Eighteenth Century," *EContentmag.com* (November 2003), <http://www.econtentmag.com/?ArticleID=5696>. Jeffrey Cymerint, The Gale Group, interview, 1 August 2003. See also Barbara Quint, "Gale Group to Digitize Most 18th-Century English-Language Books, Doubles Info Trac Holdings," *Information Today* (17 June 2002), <http://www.onlineinc.com/newsbreaks/nbo20617-3.htm>.

<sup>27</sup> *Budapest Open Access Initiative*, <http://www.soros.org/openaccess>.



# Would We Have Noticed the Loss of the Iraqi Museum?: The Case for the Virtual Duplication of Cultural Heritage Collections

Robert K. Englund

*The following lecture was presented by Robert K. Englund, Professor of Assyriology & Sumerology and Director of the Cuneiform Digital Library Initiative (CDLI) at the University of California, Los Angeles, at the National Humanities Center on October 22, 2004. As principal investigator of the CDLI, Englund has led an international group of Assyriologists, museum curators, and historians of science to make available through the Internet the form and content of cuneiform tablets dating from the beginning of writing, circa 3350 BC. Supplementing nearly half a million discrete inscriptions, translated into English and Arabic, will be online tools necessary for textual analysis allowing far greater insight into the origins of culture in the cradle of the Tigris and Euphrates rivers that can be readily accessed by scholars all over the world.*



Let's start with a snapshot of what life was like in Mesopotamia in the first millennium BC, one that might point to some parallels between what ancient Babylonians and Assyrians were facing then, and what modern Iraqis are facing today. A king of ancient Iraq, certainly someone who would have felt quite at home in one of Saddam Hussein's many palaces, wrote in the report of his first campaign against the tribes surrounding Assyria,

I massacred many of them and carried off captives, possessions, and oxen from them. I felled 200 of their fighting men with the sword and carried off a multitude of captives like a flock of sheep. With their blood I dyed the mountain red, and the ravines and torrents of the mountains swallowed the rest of them. I razed, destroyed, and burnt their cities. And into the midst of those which none of the kings my fathers had ever approached, my warriors flew like birds. I felled 260 of their combat troops with the sword. I cut off and piled up their heads. I flayed as many nobles as had rebelled against me and draped their skins over the piles of heads. I flayed many right through my land and draped their skins over the wall of Nineveh.

I cite one of the more horrific rulers of the long history of bloodletting in the Near East because, among other duties, humanists must confront the consequences of the dogs of war, in Iraq and elsewhere, once they are freed to wreak havoc on human memory. Many will remember the scene of the consummate humanist George in Albee's

*Who's Afraid of Virginia Woolf?*, an inveterate associate professor at a New England liberal arts college who declared to his biologist guest, Nick, that it was the great dilemma of conscious historians to be relegated to understanding the motivations of agents of power—in subtext, as their court poets. So I would like here to consider what I see as some of the motivations that led to the present war in Iraq, whether the roots of the present conflict may act as a warning to us to expect more of the same in the future, and what we should undertake to, in our small way, head off some of the natural consequences of this strife.

Many today find themselves playing with dark conspiracies in their attempts to understand why a grand coalition in Washington led this country to war against an enfeebled dictatorial regime in Baghdad. The American president has declared the twenty-first century a century of war on terrorism, of which Afghanistan and Iraq are the first skirmishes. Whatever you might make of his politics, and, despite some rumblings among grassroots organizations, that of a growing majority of leaders within both U.S. parties, I do believe we must take seriously the intention of this nation to project and employ its military power to preemptively thwart any threat against its vital national interest, be that threat real or perceived.

We hear much these days about a new and better future, but I think, to be realistic, we must hope for the best and plan for the worst. We in the field of Assyriology would not have wished for the need to address these issues as they apply to Iraq, but we must deal with them in a serious way, aware that we have a special responsibility to make and keep available to our peers and to our descendants the records of a civilization that, though long vanished, left so many visible traces in our intellectual and technical history. I have entitled this paper, somewhat provocatively, “Would we have noticed the loss of the Iraq Museum?” thinking above all about the level of documentation of major cultural heritage collections throughout the world, but focusing on where they are most specifically threatened, that is, in regions of great conflict.

In March and April of 2003, U.S. forces moving north from Kuwait defeated a ragtag Iraqi army, consisting of a bloated corps of well-paid officers, of diehard Saddam loyalists,

and of forced conscripts from poor villages throughout Iraq—these latter young Iraqis who took the brunt of the lethal attack were, by the way, the only participants in this war who had no choice in what was happening to them. In the midst of this invasion, there was an act of national liberation that played out on the 9th of April, 2003, at Firdos Square in the center of Baghdad—and next to, by the way, the famous Palestine and Sheraton Hotels so much in the news in the years following the invasion—with the technical assistance and under the watchful eye of a new occupation force. The footage of Saddam’s statue being toppled in that square by a U.S. tank, fed through the broadcasts of a very eager U.S. media, fairly saturated cable and nightly news reports in our homes. But beginning just one day later, a 48-hour confrontation among local Iraqis and combatants from U.S. and Iraq forces took place before and within the Iraq National Museum, itself but several miles from Firdos and its toppled statue. The confusion of war sets in with the reporting on this extended incident of cultural heritage destruction.

It appears that some of the Saddam loyalists who had scattered with the winds in the first days of the invasion had taken positions within the confines of that museum and had exchanged fire with U.S. forces. It appears further that as Donald Rumsfeld stated, “A free people are free to do bad things,” namely, that local Iraqis entered and plundered the holdings of the museum both for reasons of personal gain and out of hatred directed against the staff of that institution widely believed to have been a tool of the Baathist party. On the 12th of April, this looting was abruptly ended by the U.S. occupation force. For a short moment, the importance of preserving and disseminating cultural heritage achieved a currency among U.S. and European media and politicians that led to nervous discussions even within DoD and State Department planning staffs over how best to counter the bad publicity surrounding this apparent failure by the occupation to secure major sites of cultural heritage within Iraq.

Fully consonant with the theory of fluid group action, plunderers on the 14th of April regrouped and entered the National Library and Archives in Baghdad. They torched that unique collection, irrevocably destroying thousands of

historical documents with no facsimiles whatsoever. As in the Iraq Museum, it may be that local Iraqis saw in these archives records of the actions of a hated tyrant; we cannot know for we cannot look into their hearts. But groupthink also took hold in the circles of academicians who more closely than others followed the events in the Iraq Museum. Blogging sites hosted by the Oriental Institute of the University of Chicago, and science and culture pages of major U.S. and European newspapers came alive with reports of the mass destruction and looting of this heart of Iraqi cultural heritage. Many spoke of the complete loss of the collection, in total 180,000 unique artifacts documenting 12 millennia of human settlement, including 3,000 years of written history from the pre-Christian era. Cuneiform documents from the period of around 3300 BC until about the time of Christ were, we were told, lost for all time.

As one example, I would like to cite an article written for the *Süddeutsche Zeitung* by a professor of Assyriology at the University of Marburg in Germany: “A surprising detail in the description was the circumstance that the American soldiers themselves made the plundering possible by breaking open or unlocking well-secured gates. They then summoned bystanders to loot, saying ‘Go in, Ali Baba, it’s all yours.’ Eyewitnesses heard this standard phrase again and again. ‘Ali Baba’ had become the epitomizing term among Americans for plundering Iraqis. A witness recounted how the soldiers sat laughing on their tanks as they watched.” So now that’s the German, the European press, on what was happening in these few days of the uprising against the Iraq Museum.

These initial reports went out across the Internet, fanning fires of disgust at what was characterized as a wanton disregard for world cultural heritage by a raging band of barbarians in Iraq. The critics in these reports were not referring to the looting mobs in Baghdad and other Iraqi cities, but were pointing at those on the *American* side. As a response to this outcry, military intelligence and FBI agents were assigned the task of assessing the damage and retrieving lost artifacts, in the course of which a certain cultural propaganda war set in. While the community of specialists in museum and library science, in archaeology, and cultural history circulated and recirculated tales of damage

inflicted or tolerated by U.S. military forces in Baghdad, U.S. officials began circulating suspicions that the plunder was at least in part an inside job, since only the real pieces, the valuable pieces, were taken, and since many doors had been opened without force. Evidently feeding from some local sentiments, investigators around Matthew Bogdanos, the man put in charge of recovering Iraq Museum losses—and, by the way, the Manhattan prosecutor who grew famous with his prosecution of Puff Daddy—concluded that the Baathist museum staff must have had their own motives for stealing from their own collections.

For the record, I might state that the last time I had the opportunity to work in the Iraq Museum was in April of 1990, shortly before the Saddam invasion of Kuwait. But in the months I spent working on a specific group of cuneiform documents in that collection, I did learn that we must remain very skeptical of the description of the museum’s holdings from either or any side, since much and perhaps most of the collection was effectively undocumented. Although certainly not foreign to Western museums, the level of collection documentation within the relatively poor Near East, let alone within destitute third world countries, is truly alarming and must form a central topic for discussions among cultural heritage officials generally, and among proponents of digital libraries specifically. Clearly, we have the tools to catalogue collections quickly and at low cost, but the international community must add to this capability the will to do so. I will return to this dilemma shortly.

The list of lost artifacts has been slowly reduced by improved cataloguing and by policing work that included the use of financial incentives to pry many of the artifacts loose from their unrightful owners, a tactic that was generally supported by archiving and cultural heritage proponents in the weeks and months after the April 2003 destruction. Still, most reasonable current estimates put the loss at approximately 6,000 to 10,000 mostly small and therefore easily transportable objects—above all, cylinder seals that are a hallmark of the administrative history of Babylonia. There appears to be no image documentation of these small objects that frequent the safes of even the smallest of antiquities dealers and collectors throughout the world. A quick check on eBay in October of 2004

resulted in six probable hits of authentic cylinder seals. It is likely that the majority of these have recently been removed from Iraq.

Now, some other higher profile objects went missing, including an archaic human statue and the famous Warka Vase with its friezes of early human activity, both of which date to the end of the fourth millennium BC. A number of twenty-seventh-century statues from the Diyala region east of Baghdad, excavated by University of Chicago archaeologists in the years preceding the Second World War, were also lost. The safe return of one of these statues, which formed a centerpiece of a May 10th, 2003, exposé on the museum looting that appeared in the *LA Times*, spurred a roundtable discussion on the matter organized by the Getty Conservation Institute in Los Angeles. This discussion was attended by UCLA faculty, by Los Angeles County Museum of Art officials, and by the local community, but also by two federal investigators immediately before their departure for Baghdad. To demonstrate their solidarity with representatives of Iraqi cultural heritage, Coalition Provisional Authority officials and favored Iraqi politicians were regular guests of the reconstituted museum staff in Baghdad.

That the motives of many of the looters were unclear and often certainly unprofessional is demonstrated by the fact that most of the scenes of devastation photographed within the offices and storerooms of the Baghdad museum were the result of plunderers' intentions to steal the *furniture*, dumping stacks and piles of precious photographic and written documentation from desks and cabinets on the floors on their way out the door. I mentioned that we cannot well state with confidence how much might be missing from the Iraq Museum collections since the documentation is so unprofessional. During my own work in Baghdad, I had no immediate access to the museum storerooms, but with some regularity the curator Ahmed Kamel did bring to my table cuneiform tablets that had gotten mixed in with those that I had requested. In two instances of such unintentional largesse, I was able to make quick photographs of shoe box-sized containers of texts dating to or near the reign of Nebuchadnezzar, and thereby to underscore the desperate need for cataloguing in this and many other archives of cultural heritage across the Middle East. The unprofessional

images I made in passing are the only known record of a jumble of over 250 texts that may just document the provisioning of deported elites from Jerusalem, the prediction of solar eclipses over 500 years, or the plaintive sigh of a mother who had lost her child to a sepsis shock. Would we have noticed the loss of these texts to plunders in April of 2003? Most certainly not.

Before considering what current prospects are for the secure documentation of Near Eastern artifact collections, let me try with one example to highlight what is happening beyond the now relatively secure confines of Iraqi museums. The capture and eventual release by apparent Shiite insurgents of Micah Garen, an independent journalist from New York City, has already faded from the national media scene, but we should remember Garen as one of the real activists among proponents of cultural heritage preservation. I would invite you to bookmark his Web site at [fourcornersmedia.net](http://fourcornersmedia.net), where he and his partner, Marie-Hélène Carleton, have been documenting the widespread plunder of unprotected archaeological sites in Iraq. One image from their site exhibits some of the 1,000 cuneiform tablets recovered during a single police raid in southern Iraq in June of 2004; the quick shots made by Garen are the only photographic documentation of 1,000 relatively complete inscriptions that had shortly before their confiscation been illegally excavated at a site nearby. Such plunder often takes place with searchlights in the dead of the night, not for fear of intervention by law enforcement or occupation forces, but to avoid the deadening heat. Garen has written me that these and other artifacts were transferred to the Iraq Museum but he did not know who might be caring for or cataloguing them.

This and many other examples of countryside looting prompted University of Michigan archaeologist Henry Wright, in an edition of the *National Geographic* magazine, to rank Iraq *under U.S. occupation* as the most endangered case of cultural heritage on earth, and to worry that fifty years from now, we won't have enough of an archaeological record left to answer fundamental questions about our past and our possible future. Such matters as the guarding of significant sites of shared cultural heritage are evidently much more involved than is the relatively straightforward

issue of instituting policies geared towards the documentation and dissemination of existing collections. We might hope that law enforcement agencies will receive sufficient material support from the international community to be able to interdict the looting and cross-border transportation of cultural heritage objects wherever these crimes are taking place, particularly in an Iraq stripped of its ability to secure its own archaeological sites. But what are the prospects at least for a modest improvement of collection security in Iraq and elsewhere in the Middle East?

It must be troubling to anyone who has followed developments in Iraq in the past quarter century, and I think to those who view developments in the Middle East generally, to realize that in the United States, a nation that has, especially since the Second World War, played so prominent a role in that part of the world, no public discussion is taking place about the reason many Muslims hate us so much that they would dedicate their lives to our destruction. What really motivated those nineteen Saudis and Egyptians to commit such horrendous crimes against innocents in order to make a statement about American actions in their part of the world? It seems to me that an unprejudiced observer will look at the 2004 presidential and vice-presidential debates, let alone the national campaigns themselves, and conclude that insofar as security concerns are involved, these are highly irrational discussions. Both political parties and both candidates for the presidency seem effectively to have bought in to the argument that Islamic fundamentalists hate us because we are Americans who enjoy certain freedoms and economic and social privileges. That may in part be true, but who has conspired with the Democratic and Republican operatives to keep from public discourse the real irritants in our relations with the Middle East?

To my mind, the first is clearly our dependence on oil. This is an old point of argument that became most acute after the Arab embargo of the early 1970s. But judging from national policy on energy use since then, no federal-level legislative or executive body has proposed any serious steps to cap the profligate abuse of the world's energy reserves in this country. Energy analysts have stated that America, dependent on its own reserves, would run out of oil in a matter of several years. That is the story that we

hear regularly about once a decade, and as new reserves are found, it is as regularly pushed into the background. But I think that those who look closely at American reserves recognize that we will in fact become more and more dependent on foreign oil until such time as we institute a very different policy on energy use within this country.

We remember George Bush Sr.'s "This will not stand" proclamation before Congress prior to the Kuwait War, and his justification for that war, which was, "Most Americans know we must make sure that control of the world's oil resources does not fall into Saddam's hands." Bush Sr. was merely echoing the Carter Doctrine stating that securing Persian Gulf oil was in America's *vital national interest*, most clearly expressed in his 1980 State of the Union Address in response to the perceived threat against the Strait of Hormuz shipping lines represented by the Soviet invasion of Afghanistan. We should remember that this perceived threat to Gulf security led to the covert and overt funding of Afghani and foreign mujahideen forces, and that this same year Osama bin Laden entered Afghanistan. As the current vice president repeated in the months leading up to the effective congressional declaration of war in October of 2002, armed with weapons of terror and seated atop 10 percent of the world's oil reserves, Saddam Hussein could be expected to seek domination of the entire Middle East, take control of a great portion of the world's energy supplies, and directly threaten America's friends throughout the region. That was a speech before the Veterans of Foreign Wars in a meeting in August of 2002.

Now we compete for these same resources with new national economies that threaten to assign to a distant age the \$20 barrel of oil and the 99-cent gallon of gas. Thus oil attaches us to the Middle East in a special way. Indeed, the so-called "Bush Doctrine" presumes that the Gulf states are, in matters of national security, a part of U.S. territory, and it seems that these energy needs will conspire, with whatever party occupies the White House, to keep American soldiers stationed in the Middle East until the wells run dry.

The second point of irritation is the long-standing and often, or at least occasionally, uncritical relationship of this country to the governments of Israel. Just as images of atrocities committed against Iraqi civilians at Abu Ghraib served

to recruit new martyrs to the horrifying cause of terrorists worldwide, so too does the unresolved record of occupation of the West Bank and Gaza (and the deadly reactions and counterreactions to that occupation) commit this country to a long-term role of adversary to Arab nationalist and Islamic fundamentalist agitation in the Near East, and, as we have witnessed, throughout the world.

I mention this only by way of pointing to the situation we face as humanists who bear some responsibility for the preservation and dissemination of shared cultural heritage. We must assume that armed conflict in the form of civil war (as seems the likeliest outcome of our adventure in Iraq), cross-border hostilities sparked by nationalist fervor, a catastrophic event involving Israeli security, or an intervention by the United States or its surrogates to stabilize situations that could threaten the free flow of oil, are only some of the events that might challenge the goals of U.S. *Realpolitik* in the Middle East for the foreseeable future. What solutions might we imagine for this long-term dilemma?

Of course we could first follow imperial precedent and simply take everything to Berlin or to Chicago or London and never, ever return it, but aside from the fact that this is no longer a viable option in the modern world, the example of Berlin is a good one to warn against the idea that the West will better care for the security of shared cultural heritage than the Middle East can. Adam Falkenstein, the great Heidelberg Assyriologist, lost his extensive library to British bombing raids in Berlin, the same raids that claimed the Berlin Halaf Museum and major parts of the collection of the Pergamon that is today still being slowly reconstituted.

Failing a nationally organized removal of Near Eastern collections that so successfully filled the coffers and exhibition halls of the British Museum and the Louvre in the nineteenth century, we might hope that such international cultural and policing agencies as the FBI, Interpol, the UN, or its cultural arm at UNESCO, might play a more meaningful role in enforcing existing statutes set in place to protect national cultural heritage collections. UNESCO's 1954 "Convention for the Protection of Cultural Property in the Event of Armed Conflict," given sufficient enforcement power and given respect by members of the Security Council in New York, should form the basis for cultural heritage

protection in times of war. That respect could be signaled by the formal signing of this convention by the two main belligerents in the Iraq conflict, that is, by the United States and by Great Britain.

But on the other hand, academics and archivists who closely monitor the integrity of cultural heritage collections might cite this convention as a justification for collaboration with war planners in advance of preemptive or preventive wars. For instance, some American and British archaeologists, *before* the invasion of Iraq, met with and gave staff members of the office of Paul Wolfowitz so-called "avoidance lists" of culturally significant sites within the country that invading forces should protect from the vagaries of war—meetings which I personally find an affront to the dignity of those living Iraqis for whose homes and families such ordinance redirected from museums and archaeological sites would theoretically, through this tactic, be made available, but meetings which, in times of advancing moral relativism, were widely supported in my field. Failing the empowerment of the Hague cultural heritage convention in armed combat, we can still hope that artifacts looted during the conflict will be confiscated and returned to their countries of origin according to UNESCO's convention on the means of prohibiting and preventing the illicit import, export, and transfer of ownership of cultural property, ratified in 1972 and accepted by the United States just eleven years later.

We should not leave out of the list of current threats to Middle Eastern cultural heritage collections the possibility that state organizations might decide to destroy their own national collections. Can international organizations stop, or at least disrupt, the wanton destruction of world cultural heritage committed by a sovereign state against collections or sites within its own borders? It would appear from the recent example of the havoc played, by a ruling Taliban clique run amok, upon the great Buddha statues of Bamiyan in Central Afghanistan, but also against all pre-Islamic statues in that country, that the international community is powerless and certainly unwilling to enforce the security of what we must see as an internationally shared historical record.

It is in this sense, in the very real sense of protecting our own shared heritage as cultures in historical contact, that the Cuneiform Digital Library Initiative and other

research collaborations in the humanities can, I believe, make a difference, albeit a small one. Started in the 1980s as a cooperative effort between the Free University of Berlin and the Max Planck Society to digitize and electronically parse the proto-cuneiform collections from German excavations of ancient Warka—those are collections that date from about 3300 to 3000 BC, housed at the Iraq Museum, at the University of Heidelberg, and at the then East German Vorderasiatisches Museum—the CDLI in the early 1990s expanded its scope to include all third- and fourth-millennium cuneiform collections and in recent years to include cuneiform inscriptions generally.

In addition to digitally imaged collections in Germany, France, and the United States, we have finished work on the early cuneiform tablets in the Hermitage in St. Petersburg and have begun work on the collections of the Ashmolean Museum at Oxford and the Syrian collections in Aleppo and Idlib. We employed off-the-shelf hard- and software to capture the small objects that contained cuneiform inscriptions. Our basic text documentation is described in CDLI's Web pages, beginning with a catalogue in text transliteration, that is, in a one-to-one representation in simple text of the cuneiform inscription itself in machine-readable Roman script, and a 300 and then a 600 dpi full representation of the physical object.

While we are hopeful that such projects as the NSF-funded Digital Hammurabi effort at Johns Hopkins will eventually lead to the development of an inexpensive and easily portable 3D scanner, and browser plug-in software that will facilitate the Web dissemination of high-resolution 3D images, we are satisfied that our solution to tablet imaging—which we compare to Peace Corps efforts to develop, for instance, simple ovens that will actually continue to work for villages in Africa once Western activists have left—is currently the best answer to the needs of a community of collections that range from the private mantelpiece group of three old Babylonian letters in Fort Lauderdale, to the fifty inscriptions in the anthropological museum of the University of São Paulo, to the 100,000 pieces in the archaeological museum in Istanbul or in the Iraq Museum in Baghdad.

A more important contribution of the CDLI to the preservation and dissemination of cuneiform collections,

and we think of collections of inscriptions of dead languages generally, is the development and implementation of Extensible Markup Language description of our text corpora. In this, above all, Stephen Tinney, professor of Sumerology at the University of Pennsylvania and director of the NEH-funded Sumerian Dictionary Project, and programming collaborators working with our Berlin partners at the Max Planck Institute for the History of Science, have played a leading role. We are currently in the process of editing substantial text files to produce a consistent data set that will serve as the basis for testing our data linguistically and semiotically. The point about this descriptive means of writing up in simple text format the most important characteristics of the text is that we are writing in a language that other computer projects understand and can communicate with and exploit directly, without further input from our own team. So the idea of communication is paramount in setting up a system that is run according to XML that will make our data available everywhere today but also should put it into a form that will be easily used by generations of researchers to come.

Our cleansed transliteration files consist of over one million lines of text. This text description can now be exploited in a number of ways. CDLI's Document Type Definition (DTD) contains the description of how we code cuneiform texts in a form that is generally understandable to any other text processing research team—and indeed should be understandable with little effort to a visitor from a later age, or a distant galaxy. We have in this kind of coding chosen a path of low resistance in deciding to tag our texts strictly at the graphemic level; text structural description has been put in automatically by our XML parser to delimit what we understand to be a discrete graph. Much as with earlier instantiations of various LISP programs, our XML parser strings information in open-close structures from highest to lowest levels of text description.

So we have kept text description at this stage exceedingly simple, and have not burdened it with a lot of tagging that would describe, for instance, the meaning of the words and so on, that we have in these texts. That sort of overlay processing we leave for a later stage of our work.

Another example of how CDLI text description can now be exploited can be seen in our transliterations of archaic Persian texts, dating to circa 3000 BC. The so-called proto-Elamite texts have not been identified linguistically, yet contain sufficiently long strings of signs and sign combinations that we feel confident a computer-assisted graphotactical analysis—that is, an analysis that looks for particular strings of signs, where signs appear in longer sequences and so on—will help us to theorize about their meaning within the text. We can isolate various kinds of graphotactical strings in the full corpus, resulting, we hope, in meaningful data for at least a language typology categorization, if not a language identification of the scribes of these early texts.

As my Berlin partner Peter Damerow and I have demonstrated, using an automatic parser of our transliterations of the earliest Babylonian texts from the period slightly before that of the proto-Elamite texts, valuable statistical numerical information can be derived from multiple sign combinations—information that while probably linguistically neutral, still offers the prospect of making important semantic connections between quantitative signs in our early administrative documents, and signs that represent objects, persons, and institutions, and possibly verbal forms.

These are then the data that we gather and archive in the digital capture of a collection of cuneiform texts. With grants from the National Endowment for the Humanities and, in cooperation with the Baltimore Walters Art Museum and the Learning Federation of the Federation of American Scientists, from the Institute for Museum and Library Services, we have been developing online linguistic tools to facilitate interpretation of these texts and text archives for all user levels, including for instance lexemic or word data-mining tools for linguistic and historical research.

It seems obvious that these archived and online resources represent an important milestone in the attempt to provide cultural heritage institution officials with a reliable facsimile of their own collections in a form that is easy to use, and to scale up and down as advancing technologies make possible an improved digital capture. Metadata description—that is, files that describe the files that you have—tags all text and image files for an archival access

system that, in compatible form, is immediately available to collection managers who can build digital facsimiles of complete artifact collections. These data are fed into the communication lines of a networked international community of users that for the first time enjoys access to collections at a great and therefore prohibitively expensive distance from their home or office workstations.

The digital facsimile of physical artifacts represents our best safeguard against the many forms of expected—that is, for instance, decay of ancient objects once they are removed from their ancient strata—and unexpected artifact disturbance which we have witnessed threatening current collections. But it is obvious that these digital facsimiles can and must be expected to do more. This was, for instance, used successfully in recovering for the Iraq Museum a tablet transferred from Baghdad to a provincial Iraqi museum before the Kuwait War, and sold to a collector in London shortly after the Shia revolt in the south of Iraq. We are now developing tutorials in automatic text markers to assist law enforcement officials at distant borders, airports, or police stations, in identifying and confiscating cuneiform artifacts being stolen now. We entertain a vision that with the added urgency of stopping the flow of recent removals from Iraqi sites, international policing agencies and national and international cultural heritage statutes will institute a strict system of proof of ownership that licenses the possession of Near Eastern antiquities through a central database capture, and therefore foresees a positive ID of the pedigree of such artifacts by owners rather than by countries of origin.

CDLI text identifiers can quickly identify and track the ownership of cuneiform tablets moving through the sites of eBay, Christie's, Sotheby's, and so on, and make this information freely available through our Web pages. We of course offer our full cooperation to the International Council of Museums and to UNESCO in formatting our files for inclusion in a general database on Iraqi stolen property.

The limited cultural heritage preservation goals of the CDLI form a part of such European initiatives—spearheaded by the Max Planck Society—as European Cultural Heritage Online (ECHO). We fully subscribe to their October 2003 Berlin Declaration, stating that “in order to realize the vision

of a global and accessible representation of knowledge, the future Web has to be sustainable, interactive and transparent. Content and software tools must be openly accessible and compatible.”

The case of Iraq presents humanistic scholarship and information technology with a test. In a network that, among other tasks, serves the public mission of disseminating shared world culture—and, by the way, a network realized for the most part with public funding and using public bandwidth—can we overcome the many burdens of curatorial jealousy, of academic pettiness, of institutional and intellectual copyright, to create and disseminate intellectual and cultural content to the heirs of world culture in the United States as well as in Iraq and elsewhere? The CDLI is a modest player in this game, still one that through collaborative efforts across borders can act as a good example of cooperation in the public interest. I am therefore particularly grateful for the support that the Lyman Board and the National Humanities Center have shown our work.



# New Methods for Humanities Research

John Unsworth

*Since 2003 John Unsworth has been Dean and Professor at the Graduate School of Library and Information Science at the University of Illinois, Urbana-Champaign. From 1993 to 2003, he served as the first Director of the Institute for Advanced Technology in the Humanities, and a faculty member in the English Department, at the University of Virginia. He also chaired the national commission that produced Our Cultural Commonwealth, the 2006 report on Cyberinfrastructure for Humanities and Social Science, on behalf of the American Council of Learned Societies. He has published widely on the topic of electronic scholarship, as well as co-directing one of nine national partnerships in the Library of Congress's National Digital Information Infrastructure Preservation Program. In 2007, Unsworth secured two major technology grants from the Mellon Foundation to lead multi-institutional projects in the digital humanities. The following talk was presented at the National Humanities Center on November 11, 2005.*



Thanks to the National Humanities Center for hosting this event and the Richard W. Lyman Award, to the Lyman Award selection committee for honoring me with the award, and to the Rockefeller Foundation for funding it. It's a pleasure to be here, not least because I began my academic career at NC State, and I have many friends there, and at Chapel Hill and Duke as well.

A few years back, when I was a member of the English Department at the University of Virginia, I was talking to a nonacademic neighbor about doing research: "What?" he asked, "You discover new words?" Of course, my neighbor was thinking of "research" as something "scientific"—and had he consulted the OED, it would have supported him with a definition of research as "a search or investigation directed to the discovery of some fact by careful consideration or study of a subject; a course of critical or scientific inquiry." To tell the truth, even my colleagues in the English Department would probably have been more comfortable, on the whole, with "criticism" or "scholarship" than with "research" as the label for what we did when we were not teaching or doing committee work, unless one were talking specifically about "archival research" or possibly "library research."

Explaining how words like "method" and "research" apply to the humanities requires some retrospection, so before we look to the "*new* methods for humanities research," allow me to look backward for a moment,

first twenty years or so, to when I was in graduate school, and then another thirty years or so before that, when post-war graduate education—with its professionalization of literary study—was taking shape, in the shadow of big science.

In the 1980s, when I was going through graduate school, we certainly didn't consider ourselves to be engaged in any sort of "investigation directed to the discovery of...fact"—perhaps in history one might undertake the discovery of fact, or possibly in textual editing, but certainly not in literary criticism, not in a postmodern era. The very concept of "fact" itself had been pretty well debunked, for purposes of literary study, by the '80s, and science was simply another ideology in the service of the state. Stanley Fish and Wolfgang Iser had reauthorized the affective fallacy, Derrida's *Grammatology* taught us that in writing there was "nothing outside the text," Lyotard recommended "incredulity toward metanarratives," and the decentered self had resigned itself to the endless deferral of truth, in the desert of the real.

The other word in my title, "method," raises some issues of its own. A method is a procedure, or sometimes more specifically (as in French) a "system of classification, [a] disposition of materials according to a plan or design" (OED). In the 1980s, in graduate school (and in job interviews), one sometimes faced the daunting question "What's your methodology?" Usually, what that meant was "What's your theoretical bent: what theoretical flag do you fly?" There was an older sense of methodology still in force, though: dissertations still sometimes had chapters on methodology, and graduate programs in English were wrestling with whether or not to discard requirements for coursework in research methods (which essentially meant bibliography, sometimes with library research methods included). Most departments eventually did do away with this requirement, and by the 1990s, "research" seemed to happen mostly without attention to method.

Yet research and method are connected, logically, because systematic and organized research proceeds according to some method, some plan or design. In his 1951 *Kenyon Review* essay called "The Archetypes of Literature," Northrop Frye talked about this:

Every organized body of knowledge can be learned progressively; and experience shows that there is also something progressive about the learning of literature.... Physics is an organized body of knowledge about nature, and a student of it says that he is learning physics, not that he is learning nature. Art, like nature, is the subject of a systematic study, and has to be distinguished from the study itself, which is criticism.... So, while no one expects literature itself to behave like a science, there is surely no reason why criticism, as a systematic and organized study, should not be, at least partly, a science.... Criticism deals with the arts and may well be something of an art itself, but it does not follow that it must be unsystematic.

Systematic (methodical) thinking was, in Frye's view, what separated criticism from commentary: "commentators have little sense, unlike researchers, of being contained within some sort of scientific discipline: they are chiefly engaged, in the words of the gospel hymn, in brightening the corner where they are."

For Frye, the use of a method in pursuit of progress toward a goal was also what separated meaningful criticism from "the literary chitchat which makes the reputations of poets boom and crash in an imaginary stock-exchange.... That wealthy investor, Mr. Eliot, after dumping Milton on the market, is now buying him again; Donne has probably reached his peak and will begin to taper off; Tennyson may be in for a slight flutter but the Shelley stocks are still bearish. This sort of thing cannot be part of any systematic study," Frye maintained, "for a systematic study can only progress: whatever dithers or vacillates or reacts is merely leisure-class conversation."<sup>1</sup>

Frye's enthusiasm for the systematic, though, is probably responsible for the downturn in his own critical fortunes during the later decades of the twentieth century, when the fashion in literary criticism favored paradox, metaphor, and (in spite of the systematic basis of poststructuralism) a fairly high level of idiosyncrasy and the foregrounding of persona over logic. Oscar Wilde, who thought criticism was the only civilized form of autobiography, might have approved: Frye did not, and in 1984, in a *PMLA* essay called "Literary and Linguistic Scholarship in a Postliterate World," he remarked

disparagingly that “it has...become generally accepted that criticism is not a parasitic growth on literature but a special form of literary language.” In the end, for Frye, the insistence on the primacy of method obscured the real goal of criticism (to be “interested in literature itself and in what it does or can do for people”)<sup>2</sup> and methodology, turned in on itself, became part of the problem. In that 1984 essay, he wrote that “critical theory today has relapsed into a confused and claustrophobic battle of methodologies, where, as in Fortinbras’ campaign in Hamlet, the ground fought over is hardly big enough to hold the contending armies.”<sup>3</sup>

1984 was perhaps a low-point for both “research” and “method” in the humanities, but research did survive—perhaps perpetuated as some kind of guilty pleasure—and today it takes place quite openly, here at the National Humanities Center, as attested by this description of the Center’s activities, on its Web site:

The Center annually admits forty fellows, who represent a broad range of ages, disciplines, and home institutions. Individually, the fellows pursue their own research and writing. Together, they create a stimulating community of intellectual discourse. Interdisciplinary seminars on topics of mutual interest provide a context in which fellows share fresh insights and thoughtful criticism. The most tangible result of the fellows’ work is the publication of nearly a thousand books since the Center opened.

That’s probably a fairly good use-example of the term “research” as it now applies, and has usually applied, in the humanities: it refers to the work of an individual, work that is preparatory to writing, work that results in the publication of a book. Researchers may gather to share insights and critique, but research itself is a solitary enterprise: as the same Web page goes on to say, “Each fellow has a private study, appropriately furnished for reading, writing, and reflection, overlooking the surrounding woods.” Research in the humanities, then, is and has been an activity characterized by the four Rs: reading, writing, reflection, and rustication.

If these are the traditional research methods in the humanities, what will “new research methods” look like—and more importantly, why do we need them?

Perhaps in at least some cases, we need them because they offer better ways of accomplishing research goals that we have long pursued. So, for example, to stick with the Canadian critical archetype for just a moment longer, in 1989 Northrop Frye delivered a plenary address to a humanities computing conference in Toronto. According to Willard McCarty, who was there, Frye said that “if he were starting out to write *Anatomy of Criticism* now he would pay very close attention to computer modeling in pursuit of the ‘recurring conventional units’ of literature on which his life-work was based.”<sup>4</sup> Frye was probably a little optimistic about what computers could have done in 1989, but I think today we could actually deliver on the promise he recognized, and I’d like to spend the rest of this talk considering the ways in which new methods, enabled by information technology, can support humanities research—some new kinds of research, and some very familiar kinds of research. I’ll talk about what we can now do and what we can’t, what’s end-usable, and what requires expert intervention, what notions of the humanities—and of science—inform and sometimes distort our notion of research, and where we might really need to concentrate future graduate training, standards development, and tool-building, in order to realize the promise of these new methods for the core activities and future prospects of humanities research.

In the sciences generally, research is basic or applied. Basic research is motivated by curiosity rather than by a particular goal, and its outcomes tend to be theoretical rather than practical. Applied research usually grows out of basic research, and it usually has practical goals in view from the beginning. In reality, of course, the division between these two is not so neat, and much research could be described as one or the other, depending on the circumstances (in other words, depending on what’s being funded).

If we consider humanities research in terms of the basic and the applied, some would say that all humanities research is basic research, because it never aims at having a practical application in the sense that, say, laboratory research on transistors in the 1940s aimed at building amplifiers for electrical signals. On the other hand, if understanding is a practical outcome, then you might just

as easily argue that all humanities research is applied, in that it aims directly at producing a practical outcome, namely, changing the way we understand that part of the human record it has in view. Probably the truth is that in the humanities, as in science, both are done. Frye's work on literary archetypes, or Freud's work on the human psyche, or Saussure's work on language, might best be considered basic research: this research is aimed at developing theoretical frameworks, rather than at applying those frameworks to particular objects of attention—even though particular objects are always in view as the theories are developed. In that sense, when we apply those theoretical frameworks to the understanding of a particular text, to illuminate the text rather than to alter or extend the theory, we're doing applied research. And again, of course, in the humanities as in science, we never really do only one or the other.

In a recent instant message exchange with Steve Ramsay, a colleague at the University of Georgia who is working on the nora project (about which more in a moment), he asked:

What is a literary-critical "problem"? How is it different from a scientific "problem"? Consider the following scenario. Let's suppose that the NSF were to ask its funded physicists to report the achievements in physics for a given year. You can imagine what that list might look like. "We discovered the top quark. We achieved cold-fusion. We proved the existence of the Bose-Einstein Condensate." What if the NEH were to ask its literary critics the same question?

"Well," I argued, "that's because literary-critical 'problems' are not for solving. The object of the literary researcher is not to settle questions, but to open and explore them, whatever their rhetoric says to the contrary." Steve's response to that was that "Words like 'problem,' 'experiment,' 'fact,' 'truth,' and 'hypothesis' all mean something very different in a humanistic context than they do in the sciences." I replied: "I think we imagine science as being more scientific than it is." And I do think that (and as Steve pointed out), so did Kuhn, Feyerabend, and Lyotard. In science, one doesn't prove a hypothesis, any more than one does in cultural studies. All you can do is offer a

hypothesis that withstands being disproven, for some period of time, until contradictory evidence or a better account of the evidence comes along. For that matter, in *Against Method*, Feyerabend argued that whatever scientists might say about their adherence to methodological rules, there are no rules that they always use, and if they did adhere strictly to such rules, it would retard scientific progress. Scientific research—and the shift in ground truth during scientific revolutions—do, however, turn on evidence in a way that humanities research often does not, and science's self-correcting mechanisms are not so obviously present in the humanities. Still, the difference between scientific research and humanities research, between scientific methods and humanistic methods, may be a difference of degree rather than of kind.

Bill Wulf, president of the National Academy of Engineering, would agree, at least in the case of computer science. Bill has argued (in my hearing) that computer science should really be considered one of the humanities, since the humanities deal with artifacts produced by human beings, and computers (and their software) are artifacts produced by human beings. Harold Abelson, a professor of computer science at MIT, tells students in his CS 101 course (Structure and Interpretation of Computer Programs) that

"computer science" is not a science and...its significance has little to do with computers. The computer revolution is a revolution in the way we think and in the way we express what we think. The essence of this change is the emergence of what might best be called procedural epistemology—the study of the structure of knowledge from an imperative point of view, as opposed to the more declarative point of view taken by classical mathematical subjects. Mathematics provides a framework for dealing precisely with notions of "what is." Computation provides a framework for dealing precisely with notions of "how to."<sup>5</sup>

In other words, computers are all about method, they are epistemological to the core, and they are made by human beings. All of these qualities make them objects as well as instruments of interpretation—a point that I'll return to, after we look at some of the ways these artifacts of procedural epistemology can be used in humanities research.

My first example of new research methods for the humanities—in fact, my first several examples—comes out of the nora project. Nora (which either refers to a character in a William Gibson novel, or is an acronym for “No One Remembers Acronyms,” depending on who in the project you ask), is a two-year project funded (as so much work in digital libraries and digital humanities has been) by the Andrew W. Mellon Foundation. The project began last October, so we’re about one year in, and although I’m not quite ready to show, tonight, there is a good deal already to tell. The goal of the nora project is to produce text-mining software for discovering, visualizing, and exploring significant patterns across large collections of full-text humanities resources from existing digital libraries and scholarly projects.

In search-and-retrieval, we bring specific queries to collections of text and get back (more or less useful) answers to those queries; by contrast, the goal of data-mining (including text-mining) is to produce new knowledge by exposing similarities or differences, clustering or dispersal, co-occurrence and trends. Over the last decade, many millions of dollars have been invested in creating digital library collections. At this point, terabytes of full-text humanities resources are publicly available on the Web. Those collections, dispersed across many different institutions, are large enough and rich enough to provide an excellent opportunity for text-mining, and we believe that Web-based text-mining tools will make those collections significantly more useful, more informative, and more rewarding for research and teaching. In this effort, we are building on data-mining expertise at the University of Illinois Graduate School of Library and Information Science and on several years of software development work that has already been done in Michael Welge’s Automated Learning Group at the University of Illinois National Center for Supercomputing Applications (NCSA), developing the D2K (Data to Knowledge) software, a kind of visual programming environment for building data-mining applications.

In order to assemble the testbed for the text-mining tool development, we have negotiated agreements with a number of individual libraries, projects, and centers that hold large collections of full-text humanities resources. Our

agreements aim at producing an aggregation that has some scholarly, intellectual, and subject coherence, and they focus on nineteenth-century British and American literary texts that have been generously contributed by libraries at the University of North Carolina at Chapel Hill, the University of Virginia, the University of California at Davis, the University of Michigan, the University of Indiana, and the Library of Congress. Other contributors include the Brown University Women Writers Project, the Perseus Project, and scholarly projects at the University of Virginia Institute for Advanced Technology in the Humanities, including those on Whitman, Dickinson, Stowe, Rossetti, and Blake. These agreements have allowed us to create a testbed of about 10,000 literary texts in English, roughly about 5 GB of machine-readable text, almost all of it marked up according to the Text Encoding Initiative Guidelines. This is a small amount of data, by comparison to what’s out there in digital libraries, but it is large enough to be a meaningful testbed, and it does meet the minimum requirements for intellectual coherence.

This is a profoundly collaborative project, and very different from the solitary work that we were talking about before as being the norm in humanities research. The participants are from four universities in addition to Illinois, each site with multiple individuals, and in most cases, multiple disciplines represented as well. At Illinois, where the focus is on the data-mining itself, the work is done by two highly competent graduate students in Library and Information Science, Bei Yu and Xin Xiang, who politely allow me to muddy the waters in weekly meetings, and then proceed to get sensible things accomplished in spite of that. They also work with Loretta Auvil and others at NCSA, where the focus is properly described as engineering and applied computer science. At Maryland, Matt Kirschenbaum and Martha Nell Smith, from the Department of English and the Maryland Institute for Technology in the Humanities, and Catherine Plaisant, a computer scientist at the Human Computer Interaction Lab, work with another great group of students, including Tanya Clement and Greg Lord in English and James Rose in Computer Science. Their work focuses on visualization, and Stan Ruecker, recently added to the project from the University of Alberta, works on interface

design, along with one of his students, Ximena Rossello. Tom Horton, from Computer Science at the University of Virginia, works on software design and overall architecture, with staff at the Institute for Advanced Technology in the Humanities and with graduate students Kristen Taylor (English) and Ben Taitelbaum (Computer Science). Finally, at the University of Georgia, Steve Ramsay (faculty in English) works with his graduate student Sara Steger on developing Tamarind, the XML data-management system that supports the project's need to query large XML collections for quantitative information in real time.

That's seventeen people, on one project: seven faculty members, one NCSA staff person, and nine graduate students. And I've probably left someone out. The project is divided up fairly neatly, into data-mining, interface/ visualization, data support, and architecture, but it is a real challenge to do this kind of thing, on many levels. First, simply coordinating lots of people is difficult. I think we had a breakthrough on that front when we arrived at the point where the tasks were sufficiently well defined and the goals sufficiently clear that faculty could get out of the way and let graduate students work directly with each other. Second, we're building something that none of us (or anyone else) has ever seen before, so a large part of the problem is figuring out exactly what it is supposed to be and how it is supposed to work. Third, each time we try something new, it has ramifications across the whole system, and sometimes that means that we have to stop and tear something apart and rebuild it, before we can move on to the next step.

There are many more challenges than I'll mention tonight, but perhaps the greatest challenge, at the outset and still today, has been in figuring out exactly what data-mining really has to offer literary research, at a level more specific than the cleverly nonspecific generalities I offered in my opening description of nora ("software for discovering, visualizing, and exploring significant patterns across large collections of full-text humanities resources"). What patterns would be of interest to literary scholars? Can we distinguish between patterns that are, for example, characteristic of the English language, and those that are characteristic of a particular author, work, topic, or time? Can we extract patterns that are based in things like plot, or syntax?

Or can we just find patterns of words? When is a correlation meaningful, and when is it coincidental? What does it *mean* to be "coincidental"? How do we train software to focus on the features that are of interest to researchers, and can that training interface be usable for people who don't like numbers and do like to read? Can we structure an interface that is sufficiently generalized that it can accommodate interest in many different kinds of features, without knowing in advance what they will be? What are meaningful visualizations, and how do we allow them to instruct their users on their use, while provoking an appropriate suspicion of what they appear to convey? How would we evaluate the effectiveness of our visualizations, or the software in general? Is it succeeding if it surprises us with its results, or if it doesn't? How can we make visualizations function as interfaces, in an iterative process that allows the user to explore and tinker? And how in the *hell* can we do all this in real time on the Web, when a modest subset of our collection, like the novels of a single author, contains millions of datapoints, all of which need to be sifted for these patterns? It takes many different kinds of expertise, and many hands, even to bring the epistemological elements of all this into focus, and as much again to work out the procedural details involved in actually building something that would allow a researcher to look for patterns across large collections.

As a part of the nora experiment, we're going to try to use text-analysis techniques to answer some of these questions—if not empirically, at least with some combination of evidence and subjective analysis. To that end, at my suggestion, Bei Yu has been analyzing literary criticism (from online journals in Project MUSE) and comparing it to normal usage (as represented in the American National Corpus, a just-released collection of about 10 million words from the *New York Times*, *Slate*, and other such sources), and to journal and conference literature from the knowledge discovery domain (in other words, from data-mining). What we're looking for are words common in literary criticism *and* data-mining, but not common in the *New York Times*. The theory is that this will provide at least a start on figuring out an answer to the question "What do literary scholars already do, that data-mining can support?" So far, there have been some interesting results.

In the first stage of this research, Bei generated lists of relatively unusual verbs from MUSE journal articles. She then asked a literary scholar (me) to identify some that seemed to indicate critical behaviors that might be characteristic of literary scholarship. Obliging, I did so, and picked out words like “destabilizes, annotates, juxtapose, evaluates.” Then she ran this list against the American National Corpus, and found that none of the words I’d picked were actually unique to literary criticism or even much more common in literary criticism than in normal usage. However, comparing her whole set of journal articles to the ANC, she found quite a few verbs that *were* unique—for example, “narrating, obviate, misreading, desiring, totalizing, mediating.” Her conclusion, concerning round one?

Actually, the verbs...picked out by the literary scholar turn out to be common in ANC-NYTIMES corpus too. However, after examining the unique MUSE verb list, two literary scholars were surprised to find many unexpected unique verbs, which means their uniqueness is beyond the scholars’ awareness. In conclusion, *literary scholars are not explicitly aware of what are the unique research behaviors at the vocabulary-use level.* They might be able to summarize their scholarly primitives as Unsworth did.... But this does not help the computer scientist to understand the data-mining needs in literary criticism.

This *lack of explicit awareness* on the part of the critic will become a leitmotif as we continue to discuss text-mining in literary contexts, so let me flag it as it arises here for the first time.

Bei also tried topic-analysis of the MUSE articles, to see if that would help turn up some things for data-miners to do. She found

that many essays are trying to build connections between writers, characters, concepts, and social and historic backgrounds. As evidence, 56 out of 84 ELH essays and 24 out of 40 ALH essays titles contain “and”—one of the parallel structure indicators. For example:

\* *“Monumental Inscriptions”: Language, Rights, the Nation in Coleridge and Horne Tooke*

\* *“Sublimation Strange”: Allegory and Authority in Bleak House*

\* *“Tranced Griefs”: Melville’s Pierre and the Origins of the Gothic*

\* *Passion and Love: Anacreontic Song and the Roots of Romantic Lyric*

In conclusion, simple MUSE topic analysis does not help to find new data mining applications. The reason might be that topics are so high-level and abstract that they cannot be easily represented as countable lower-level linguistic features for data mining purposes.

The third step in this procedure, or method, was then to compare the journal literature from data-mining with that from literary criticism, and see what words, at least in our sample, seemed to occur frequently in both, but infrequently in the American National Corpus. Some of those words were “model, pattern, framework, spatial,” and various forms of the words “classify, correlate, associate, relationship, similarity, hierarchy,” and “sequence.”

The final step is to sit down with literary scholars and look at the phrase-level context for these words in the criticism itself, to see where these words—representing interests that seem to overlap between data-mining and literary criticism—actually refer to things that data-mining could support in literary criticism. For example, since pattern is our declared objective, here are some of the phrases in which one finds the word “pattern” embedded, in literary criticism:

- patterns of manuscripts
- narrative patterns
- gender-inflected patterns of viewing status of women
- rhetorical patterns of gendered voices
- patterns for formation of identity
- consumption patterns
- urban-based cultural patterns
- daily patterns of intimacy, work, and play
- metrical patterns: meter and rhyme
- patterns of relationship
- marriage patterns
- patterns of plot

Which of these patterns can we actually find in large collections? Well, we went looking for some, to find out. We began with a pattern suggested by Matt Kirschenbaum, namely, the use of erotic language in the poetry of Emily Dickinson. The object would be to have Dickinson scholars identify erotic and nonerotic Dickinson poems (hot and not hot, for short) and the vocabulary that makes them so, and then subject the same corpus to analysis by software, to see if the software, trained by the expert judgments, could learn to predict which poems would be hot, which not, and why. Early on, Matt worried that the net of all this, if we were successful, might be

that our results may...largely confirm information the scholar already has in hand, or at least strongly suspects. While I hold out hope that our visualizations will contain a genuine surprise or two, there's a larger sense in which they'll merely be confirming (or suggesting) what we already know: that a high "hotness metric" for a given document suggests that that document is likely to be of interest to the scholar who supplied the particular indicators in the first place. This isn't as circular as it sounds (the computer is working on a scale and at a pace that would be impractical for a human investigator performing the same analysis manually) but this is still a very traditional form of text analysis and does not, it seems to me, take advantage of any actual data mining algorithms. What we ultimately want to be able to do is have the computer (I use the word loosely) suggest a new indicator, one conjured automatically from the list of indicators described (based on proximity or other forms of pattern matching). The idea, in other words, is for the computer to show us that [some word] crosses some threshold in relation to its proximity to the words already listed, thereby startling some intrepid Dickinsonian (who might that be?) to stroke her chin and say, "Hmm, I wouldn't have thought of that but it sure is interesting. I'm going to go and reread some poems I thought I knew well."

A computer science graduate student in Ben Schneiderman's Information Visualization class at Maryland, Nitin Madnani, put together a Java tool for visualizing weighted searches across multiple poems, so that it would

be easy to see the poems in which erotic terminology, once identified, seemed to cluster. This was done outside of D2K, and very hard-coded. It was an experiment in visualization and its requirements in this context (and in that respect, it will inform what we do later in designing interfaces and visualizations to work with D2K), but it was also a tool that helped the literary scholars begin to explore patterns across multiple works.

Having tinkered with that tool a bit, Martha Nell Smith and Tanya Clement sat down with the corpus of Dickinson poems and labeled each text hot or not—as a whole. This is, of course, a subjective evaluation, but it also represents expert knowledge. These evaluations were passed to Bei Yu, who subjected the corpus to a kind of predictive analysis known as Naïve Bayesian classification (no relation: NB is named for Thomas Bayes, an early eighteenth-century British mathematician and minister). As Matt Kirschenbaum explained in a recent presentation on the nora project,

Bayesian probability is the domain of probability that deals with non-quantifiable events: not whether a coin will land heads or tails for instance, but rather the percentage of people who believe the coin might land on its side; also known as subjective probability. Our Bayesian classification is "naïve" because it deliberately does not consider relationships and dependencies between words we might instinctively think go together—"kiss" and "lips," for example. The algorithm merely establishes the presence or absence of one or more words, and takes their presence or absence into account when assigning a probability value to the overall text. This is the kind of thing computers are very good at, and naïve Bayes has been proven surprisingly reliable in a number of different text classification domains.

The purpose of all this was to predict what we thought we already knew, namely, what makes a Dickinson poem erotic. The prediction done by experts was based on vocabulary, but it was more generally based on long experience in writing about and reflecting on Dickinson poems—in other words, it was based on traditional humanities research methods. The prediction done by the nora software was based on the combination of the experts' overall determina-

tions, evaluated against some 4,000 features—in this case, words—extracted from the document set and ranked according to the probability that they would appear in erotic poems. In some cases, humans and software agreed. For example, the words “tasted, faces, touching, Lords, Berries, feel, Nights, Hands, Nut, Butterfly, seal, Queen, and Bees” were all identified as highly correlated with eroticism, both by the experts and by the nora software. In some cases, though, the software contradicted the experts: for example, the words “Music, tune, warm, cold, Lightning, blood, Sun, cut, and love” were all predicted as markers of the erotic by the experts, and found not to be, by the software. Most interesting, to me at least, in the list of contradictions: terms in which eroticism varies by number. For example, it is erotic to be plural in the case of “nights, bees, berries, hands,” and “faces,” but those same words, in the singular, do not register as markers of the erotic. Conversely, “nut” is erotic, but “nuts” are not. Hmm.

Finally, and most interestingly, the software turned up some highly correlated markers of the erotic in Dickinson that hadn’t appeared on the experts’ list: “mine, must, Bud, Woman, Vinnie, joy, Thee, write, Eden, luxury, remember,” and “always” followed by three dots.

Here’s what the Dickinson scholar, Martha Nell Smith, had to say about those results, in a post just this week, to nora’s e-mail list ([webviz@prairienet](mailto:webviz@prairienet)). It’s a bit long, but worth reading in its entirety:

In the 1990s Harold Love stated something very important about ‘undiscovered public knowledge’: that too often knowledge, or its elements, lies (all puns intended) like scattered pieces of a puzzle but remains unknown because its logically related parts are diffused, relationships and correlations suppressed. Five years ago I wrote about that fact in “Suppressing the Book of Susan in Emily Dickinson,” an article surprisingly few Dickinson scholars seem to know, precisely because, though it’s well situated in the volume *Epistolary Histories*, it is separated from much Dickinson criticism. Love does not remark anything we don’t already know in some way, shape, form, and that is, I suppose, precisely the point. At one point or another members of this nora team have been frustrated over the “oh wow” moment

that just seemed to be missing. When my first one came, I was left saying, “uh, duh”—the “oh wow” moment is right in front of me/us. (By the way, all of my observations here are drawn from thinking about the lists that Bei sent and came out of conversations with Tanya and Catherine.) When Bei sent the computationally-generated list of found erotic terms and “Vinnie” was a “hot” term, and one of the most frequent to occur, I was at first surprised. But just a smidgeon of reflection changed that surprise to “uh, duh” recognition. Of course I had known that many of Dickinson’s effusive expressions to Susan were penned in her early years (written when a twenty-something) when her letters were long, clearly prose, and chock-full of the daily details of life in the Dickinson household. But I had never thought of this fact in quite the way that the data mining “search and find the erotic” exercise made me put together the blending of the erotic with the domestic. And thus I was surprised again because I’ve written extensively on the blending of the erotic with the domestic, of the familial with the erotic, and so forth. So I should have expected “Vinnie” to appear frequently in these early letters and to appear near erotic expressions, but I was still taxonomizing (and rather rigidly so) in my interpretations without realizing I was doing so. In other words, I was dividing epistolary subjects within the same letter, sometimes within a sentence or two of one another, into completely separate categories, and I was doing so un-self-consciously. I could wax eloquent here about why understanding the erotic as part and parcel of, and not separate from, daily life is so important, but in the interest of time I’ll just note the important connection, a connection discouraged by the traditional hierarchies of Western culture. Making the connection leads to critical understandings not otherwise obtainable, and the data mining exercise helped me do that. Similarly, though I had not designated “mine” as a hot word, it did not surprise me at all that it was FIRST on Bei’s list. The minute I saw it, I had one of those “I knew that” moments. Besides possessiveness, “mine” connotes delving deep, plumbing, penetrating—all things we associate with the erotic at one point or another. And Emily Dickinson was, by her own accounting and metaphor, a diver who relished going for the pearls. So “mine” should have been identified as a “likely hot” word,

but has not been, oddly enough, in the extensive literature on Dickinson's desires. Same goes for "write"—oh to leave a piece of oneself with, for, the beloved. To "write" is to present oneself, or a piece of oneself, physically—and noting that the data mining was picking up both "write" when recorded by Dickinson and "write" in the [XML] header [where it would indicate a letter] led the three of us to a "can we teach a computer to recognize tone" discussion. I wonder, remembering Dickinson's "A pen has so many inflections and a voice but one," what the human machine can do, what the human machine does (recognizing, identifying tone), what we think we're doing when we're so damned sure of ourselves. So the data mining has made me plumb much more deeply into little four- and five-letter words, the function of which I thought I was already sure, and has also enabled me to expand and deepen some critical connections I've been making for the last 20 years. On this list I've already talked about the limitations of "key words," a fact of which all humanists who get frustrated with search and retrieval are all too well aware, so I won't go on at great length about that. "Key words" are indispensable, but they don't work like magic, and we need to be rigorously self-conscious about all such taxonomies. I knew that, but it still surprised me when I saw texts that had several key erotic words and the texts were definitely not "hot." So Harold Love's observation very much holds—all of this was available to me but lay scattered as unrelated pieces. The data mining exercise was key to pulling it all together. Oh, and perhaps it goes without saying that the exercises also made me pull some things apart in order to make these connections.<sup>6</sup>

To this, within an hour, Steve Ramsay replied:

What, then, is this shock of recognition we feel? How do we make sense of it? Is it useful? We're all familiar with McGann's memorable remark (from Lisa Samuels, I believe) that HC is all about "imagining what you don't know." But here, we seem to be encountering something different: imagining what we already know. And in a sense, won't data mining operations of the sort we are undertaking always produce this effect? After all, we trained the system. It only knows how to look for what is already implicit

in our sense of things. It will produce a more granular, more all-encompassing vision of what we know, but what "we know" is the ground of its knowing. I am repeatedly asked questions like "Well, who decides what the erotic is?" I say two things about that. First, no one is actually defining [eroticism]. It's much more akin to Justice Stewart's observation about the obscene ("I know it when I see it"). We don't define the term so much as point to the instances we believe belong to its class. Second, whatever deciding is going on is as highly subjective, as insistently contingent as any other critical act. The fact that we are subjecting it to computational analysis neither diminishes nor enhances the implications of that fact. But if highly subjective interpreters point to instances of a particular class, and the computer comes back with the defining features of that class, have we done anything other than give ourselves a deeper understanding of what is implicit in our own subjective musings? Much will depend on how we present that insight, I think."<sup>7</sup>

Martha clearly thinks that it is a worthy outcome to arrive at a deeper understanding of what we already know, but I think she'd also argue (and maybe she has—I haven't checked the mail today) that when the data-mining process throws up new markers of the erotic, at least some of them lead her to new understandings of Dickinson, and don't just confirm or expand the understandings she came with. Data-mining delivers a new kind of evidence into the scene of reading, writing, and reflection. Although it is not easy to figure out sensible ways of applying this new research method (new, at least, to the humanities), doing so allows us to check our sense of the gestalt against the myriad details of the text, and sometimes in that process we will find our assumptions checked and altered, almost in the way that evidence sometimes alters assumptions in science.

We're continuing on with these experiments, and the next round will take it up a notch from the works of an author to instances of a genre. We're looking at sentimental fiction in the nineteenth century. The first round of training, completed by Kristen Taylor and others at Virginia, consists of ranking each of the chapters of *Uncle Tom's Cabin* on a one-to-ten scale. A number of people do these rankings and they are compared, and again we look for markers in vocabulary. We'll then take these results to a number of other

works—ones that we recognize as sentimental and others we don't—and we'll see what we learn. So far, we already see some things that are interesting in the context of this particular novel: in the top 100 words appearing in chapters rated highly sentimental, number one, with a bullet, is "Senator" and the rest of the top ten are "susan," "weeping," "bird," "reflections," "auctioneer," "cloak," "john," "block," "mud." "Mothers" doesn't show up until #16; "forgive" is quite a bit higher on the list than "defenceless"; "pain" and "prison" beat out "agony" and "sorrow"; and way down toward the bottom of the list are words like "rose-colored," "swaying," and "melted." Writing to the e-mail list about the high ranking of "bird," Kristen said:

"bird" at 4 is cool. Most of the occurrences are in the highly sentimental chapter 9... with the Senator and Mrs. Bird, but there are enough significant usages of the word applied as an adjective (only once does it refer to actual birds) to make it significant. This would be a cool paper—Stowe is riffing off the slave song "I'll Fly Away," but the 'flying' and 'escaping' words do not appear often.

"Imagining what you already know" is a good description of modeling in many humanities contexts. For example, in building a model of Salisbury Cathedral, or the Crystal Palace, as we did at the Institute in Virginia, you could say that we were imagining what you already know about those structures. However, interestingly, the act of modeling almost always brings to the surface of awareness things you didn't know you knew, and often shows you significant gaps in your knowledge that—of course—you didn't know were there. Of course, in some cases—maybe even in all cases that I've mentioned—one could (in principle) do this kind of modeling and even the quantitative analysis without computers. You could model the crystal palace with toothpicks and plastic wrap; you could do the painstaking word-counting and frequency comparison by hand. But you wouldn't, because there are other interesting things you could do in far less time.

Near the beginning of this talk, I raised the distinction between basic and applied research. From a data-mining point of view, what we're doing in the nora project, for the

most part, is applied research. We're not developing new alternatives to naïve Bayesian analysis, for example. But from the humanities perspective, I would argue what we're doing is basic research, because we are working out research methods that can then be applied in pursuit of more immediate research goals (like developing new understanding of particular texts). There are many other new research methods, in addition to statistical analysis, that are on the horizon, too, and needing (at least from the point of view of the humanities) some basic research. Simulations, games, map-making, semantic and semiotic tools—there's a lot of this kind of work yet to do, a lot of basic work, to bring information technology to bear on humanities research. Doing that work will require interdisciplinary teams, because there's too much in any of these projects for one person to do, and because it is simply impossible that any one person would have all of the necessary expertise. The problems that these teams will encounter will, I'm sure, be substantially the same as those we've been encountering in the nora project—and perhaps the work that Bei Yu is doing will provide a reusable method for determining the best fit between the capabilities of tools developed in other domains, when they are brought to bear on research in the humanities. In that respect, what she's doing may be the most basic research of all, in spite of its focus on application.

It is easy to predict that new kinds of graduate training—at least, new for humanities graduate students—will be both necessary and available in this kind of collaborative project work. You've got to have graduate students involved, because they have so much to contribute in actually carrying out certain parts of the research program, and by the same token they can make some of those parts their own, get their own publishing done, and build dissertations out of the raw materials in something like nora. They can be funded while doing it, too, and they have a completely different kind of working relationship with faculty than that provided by the tutorial model that still informs most graduate training in English. They work with faculty in other universities, which has real significance when they hit the job market, and they work with graduate students in other universities and in other disciplines as well, which means that

they have a very different sense of community throughout their graduate careers than do most of their peers, at least in English departments.

The other thing I think I can predict with some confidence is that computational methods for humanities research require a new kind of infrastructure. We've been building the digital library for some time now, and the library has always been the research infrastructure for the humanities, but in order to support this new kind of research, digital libraries are going to have to interoperate in ways they are not really even close to doing now. And for certain kinds of things—like data-mining—it's hard to imagine being able to derive the requisite quantitative information from collections that are distributed rather than aggregated: to put it simply, you need to put things in a big pile to find out the characteristics of that pile. The good news, though, is that infrastructure is, by its nature, somewhat general purpose. You can use electricity to drive lots of different devices, and you can use something like Tamarind—Steve Ramsay's XML data management system—to answer lots of different kinds of questions. We don't have to build new infrastructure for every new project, especially if we've properly distinguished between basic and applied research. Growing out of nora, for example, I can already see a set of applied research activities—probably taking the form of journal articles, actually—and some proposal for further basic research to develop infrastructure. That latter work will focus on bringing the nora testbed of eighteenth- and nineteenth-century British and American literary texts together with earlier texts that are being similarly prepared and analyzed in a project called "WordHoard," run by Martin Mueller (in English and Classics, at Northwestern University). Taken together, the infrastructural work that's being done in these two projects can, we think, form the basis for an Analytical Corpus of English and American Literature that would support many different applications in humanities research, across many different kinds of literature and literary study.

On the subject of "infrastructure" I'd like to encourage you to have a look at the draft report of the Commission on Cyberinfrastructure for Humanities and Social Sciences, sponsored by the American Council of Learned Societies. It

became available for public comment just a few days ago, and it can be downloaded from the ACLS Web site. The commission is looking for comments on this draft, and your contributions would be most welcome. We hope that when it is complete the report will help to foster the development of the tools and the institutions that we require in order to reintegrate the human record in digital form, and make it not only practically available but also intellectually accessible to all those who might be interested in it.

That goal is, I think, a good place to stop, because it brings us back to the point that Frye made about the purpose of criticism in general, which is that it should be "interested in literature itself and in what it does or can do for people." However "scientific" or statistical or technical these new research methods might seem—however systematizing, totalizing, and Gradgrindian—they are driven by the desire to understand the human record, and perhaps even more, to understand our understanding of it. That it should take a machine to do that is only a superficial paradox. The machine itself is simply an instrument of procedural epistemology, and its only function, at least in humanities research, is to offer us methods for imagining what we don't know, as well as what we do.

## Notes

<sup>1</sup> Northrop Frye, "The Archetypes of Literature," *Kenyon Review* 13, no. 1 (1951): 92-110.

<sup>2</sup> *Spiritus Mundi: Essays on Literature, Myth, and Society* (1976), 106.

<sup>3</sup> "The MLA and Literary and Linguistic Study and Teaching: The Centennial Forum. John H. Fisher; Geoffrey Marshall; John William Ward; Helen Vendler; Richard Lloyd-Jones; Frank G. Ryder; Northrop Frye," *PMLA* 99, no. 5 (October 1984): 974-95. Frye's contribution has the title mentioned above, and these passages are from page 991.

<sup>4</sup> Willard McCarty, "Computing the Embodied Idea: Modeling in the Humanities," *Deutsche Gesellschaft für Semiotik, Universität Kassel*, July 19, 2002, <http://www.kcl.ac.uk/humanities/cch/wlm/essays/kassel/>.

<sup>5</sup> [http://mitpress.mit.edu/sicp/full-text/book/book-Z-H-7.html#%25\\_chap\\_Temp\\_4](http://mitpress.mit.edu/sicp/full-text/book/book-Z-H-7.html#%25_chap_Temp_4). See also the video of lecture 1, at <http://swiss.csail.mit.edu/classes/6.001/abelson-sussman-lectures/>. Thanks to Steve Ramsay for this reference.

<sup>6</sup> Martha Nell Smith, e-mail of 11/10/05, subject "Curmudgeon Reflections on nora," to [webviz@lists.prairienet.org](mailto:webviz@lists.prairienet.org), the project e-mail list for the nora project.

<sup>7</sup> Steve Ramsay, e-mail of 11/10/05, subject "Re: Curmudgeon Reflections on nora," to [webviz@lists.prairienet.org](mailto:webviz@lists.prairienet.org), the project e-mail list for the nora project.





# The Imaginations of Computing

Willard McCarty

*Willard McCarty, reader in humanities computing at the Centre for Computing in the Humanities, an academic department in the School of Humanities at King's College London, is best known as a theoretician of the digital humanities. In his 2005 book, Humanities Computing McCarty explains how and why humanities computing is in itself an intellectual humanistic field of inquiry. His body of work in the field includes projects such as The Analytical Onomasticon and "Humanist," a digital medium bringing together scholars working on problems at the intersection of computing and the humanities, which he founded in 1997 and has helped develop into an international digital resource for all humanities scholars with interests in humanities computing. The following lecture was presented at the National Humanities Center on November 6, 2006.*



This Lampe through all the regions of my braine,  
Where my *soule* sits, doth spread such beames of grace,  
As now, me thinks, I do distinguish plain,  
Each subtill line of her immortall face.

—Sir John Davies, *Nosce Teipsum: Of the soule  
of man and the immortality thereof* (1599)

For Northrop Frye

## I. Going wide . . . . .

The style of criticism apt for computing is the very one it enables: an indefinitely wide inquiry across the scholarly conversations of all the humanities, the nearby social sciences, and beyond. Its criticism demands breadth because computing is methodological: methods migrate, bringing with them informing remnants of their accomplishments elsewhere. In earlier days many a scholar spent hours in a library carefully taking notes on if not transcribing essential writings verbatim. We were encouraged by the constraints of the time to focus in and go deep, ranging more slowly over fewer things. Now, with old barriers down, we can go wide as never before. The dangers of doing so are real, obvious, tiresomely moralized, and largely unheeded. But Googling-for-whatever is a sign of the times that points in at least two directions. Apart from the Spenglerian, a far more interesting trajectory is plotted by Hans-Georg Gadamer's nomalism, which gives

breadth to its credentials. Richard Rorty explains Gadamer's argument in a recent commentary: "no description of an object is more true to the nature of that object than any other"—because *there is no essential nature for description to reach*. If we accept this, then "the more descriptions that are available and the more integration between these descriptions, the better is our understanding of the object" than *any* one of them could possibly yield (Rorty 2000: 23–24). Emphasis thus shifts away from absolutes to the great task of integrating contingencies—of getting others to speak one's language and learning to speak theirs. The shift is away from quiet burrowing to wide-ranging conversation.

Working out in practice how to do breadth well, on the Web, is one of our main tasks, first as researchers, then as designers and teachers. Here I will go wide in aid of a very large question I have found impossible to avoid, and which I think is impossible to approach successfully otherwise: what computing has to do with imagination. This is to ask not merely how to be imaginative with our machines, rather more how imagination and computing shape each other.

This question is too large to begin with, so allow me to start small, with one kind of imagination, the historical.

## II. Getting historical . . . . .

Recently, along with a few others in humanities computing, I have begun to worry about being without a proper history. In response, I have this autumn begun to teach an undergraduate course entitled "Readings toward a History of Humanities Computing." Its objective is to see if we can begin to move on from chronologies and timelines, even though the evidential record is fragmentary and as messy as one might expect from events well within living memory. In March 2003, I participated in a conference at MIT on the history of recent science, in preparation for which I learned that some (but definitely not all) historians think that recent history can be written.<sup>1</sup> Even so, as someone involved in building a field that has barely a purchase on the academy, my aim is not so much a recent history as a "history of the present," to use Foucault's phrase.<sup>2</sup> I am aware of Whiggish peril, but disciplinary survival, and more than that, concern

for survival of the humanities as forms of life worth living, impel me to make the attempt to "recognize and distinguish historical objects in order to illumine our own predicaments" (Hacking 2002: 182). Doing so certainly raises difficult historiographical questions, one of which will preoccupy me shortly, but it is first with the need for a historical imagination that I am concerned.

Fifty years ago, in *Anatomy of Criticism*, Northrop Frye sketched the normal path of development for the systematic field he hoped his would become—for a "science," he said, and drew sharp criticism from reviewers for suggesting that literary criticism should be one. (Let that pass for now; I will return to the sciences later.) According to his sketch, a new field begins in a state of naïve induction, its phenomena taken as axiomatic, therefore their chronology as its history. Development culminates when practitioners discover that their real function is to interpret these phenomena (1957: 15). When they are able to do that, what was their cosmology becomes their subject.

We in humanities computing linger at the naïve stage. We may be somewhat less inclined than our more technical colleagues to recite chronicles of firsts. The chronology that comes easiest to mind for us takes the form of a professional autobiography, which runs from precocious analytical childhood, through Web-besotted adolescence, to reflective if bemused adulthood. We begin with the first person working on material recognizably in the humanities, using equipment that even then was called a computer, ignoring all that came before, in all its rich complexity, and we keep the fences in place from this beginning until now. But to get a genuine history we require what Frye, paraphrasing Bacon, called "a great inductive leap" to a new vantage point from which the events we chronicle appear, in the full tapestry of their times, in altogether different shapes than chronology allows for. Shapes that confer meaning beyond the dreary "to-morrow, and to-morrow, and to-morrow."

Why should anyone care? Because without history in the proper sense a field of enquiry has no epistemic coherence, and without that it is not merely weak but little more than plunder for its neighbors—"just a tool" or resource to be used at pleasure, without essential challenge to the user's conventional practices. Nothing to say *for itself*,

therefore nothing of substance to say to anyone else. I have spent two decades working on that speech-for-itself so that humanities computing can give back by talking back. For although most practitioners operate largely in ignorance of anything beyond their disciplinary ken, each discipline exists to challenge the rest with their essential incompleteness. Each, by its challenge, offers the possibility of a leap into a larger intellectual world, where (again to quote Richard Rorty) one's own "agreed-upon set of conventions about what counts as a relevant contribution, what counts as answering a question, what counts as having a good argument for that answer or a good criticism of it" are relativized by encountering other possibilities (1980: 320). The expansion of mind on offer is analogous to what happened when in the European Age of Exploration folk-biological taxonomy encountered unclassifiable forms of life (Atran 1996/1990), or much later, when the ethnographies of participant-observing social anthropologists began loosening the moorings of a worldview so settled that it seemed the world (Geertz 2000/1983). If humanities computing can gain a historical sense of going somewhere, inflected by but different from technological progress and the shifting fashions of the academy, then it has something to give without which the humanities, and therefore *all* of us, will remain as poorly equipped to meet the world's possibilities as we have become. "For the interesting puzzle in our times," Langdon Winner has written, "is that we so willingly sleep-walk through the process of reconstituting the conditions of human existence" (1997/1986: 61).

### III. The lay of the land . . . . .

I begin with a curious problem, namely, why related events cluster in time, which is to say, how they are related other than by time. A primitive way of imagining the situation is with timelines, which are commonly used to visualize technological chronicles of firsts. They are compelling, I think, because their one-dimensional geometry suggests both a straightforward causal chain and, because it cannot be avoided, the quasi-providential authority that chooses which events to record. Not the best way to think if we are

to make best sense of our lot here below. Better is to imagine clustered events as *cognate*, and so to invoke a Wittgensteinian *family resemblance*—"a complicated network of similarities overlapping and criss-crossing" rather than a single, authoritative origin (*PI* 66–57). Or we may speak of things in time as *confluent*, and say that although gravity pulls inexorably on them as they flow along, their course is shaped by the lay of a land so complex as to escape any single explanation or ordering principle—and so to require more of an aesthetic than a logic.

Commenting on "the confluence of ideas in 1936," when Alonzo Church, Stephen Kleene, Alan Turing, and Emil Post separately proposed exact definitions of the mathematical idea from which computing later arose, Robin Gandy notes, adding yet another metaphor, "There is something in the air which different people catch" (1995/1994: 51). Toward the end of his life Kurt Gödel, who in 1931 had rescued mathematical truth from the clutches of proof, invoked the notion of a *Zeitgeist* rather than an airborne disease to denote the larger context in which he, Turing, and the others had been working (Gödel 1995/ca. 1961). This context had been largely defined by Göttingen mathematician David Hilbert's passionate drive for totalizing, formalizable knowledge. In 1900 Hilbert had laid out the agenda for his field in a famous lecture in Paris (Hilbert 1902). In 1917, as the empires of Europe were coming to a violent end, he spoke, apparently without irony, of disciplines as imperial states (Ewald 1996: 1107). But, as noted, his hopes for his own empire-of-mind were demolished years later by Gödel's rescue of truth from proof, then by Turing's demonstration that there is no general algorithm to determine whether mathematical statements are true (Turing 1936). Computing was a byproduct of that demonstration—a refuge, if you will, for the completeness and certainty Hilbert desired and that imagination—whether of mathematics, poetry, the plastic arts, or otherwise—transcends. And there, precisely there, we have computing's greatest gift. But I am getting ahead of myself.

Fast forward now to the latter years of World War II, to northern Italy, where the Jesuit scholar Roberto Busa pursued interests in philosophy and philosophical texts, as he says, "surrounded by bombings, Germans, partisans, poor

food and disasters of all sorts” (1980: 83). Fascinated by the idea of inwardness in the writings of St. Thomas Aquinas, he realized that understanding lay in the minutest study of language, and so resolved to produce the *Index Thomisticus*, “a concordance of all the words of Thomas Aquinas, including conjunctions, prepositions and pronouns.” The idea of such an index was not new, but—here is Busa’s crucial realization in his own words— “It was clear to me that to process texts containing more than ten million words, *I had to look for some type of machinery.*” Look he did. In 1949 he flew with an assistant across the Atlantic to visit twenty-five or so American universities, ending up in the office of Thomas J. Watson Sr., head of IBM, whom he persuaded to donate the necessary equipment. He did this with such force that Watson’s only stipulation was “that you do not change IBM into International Busa Machines.”

Consider: Is it even remotely reasonable to ascribe such conviction and effort merely to what in war-torn northern Italy could have been no more than a rumor of suitable machinery? Do we solve (that is, dismiss) the problem by referring to genius, placing Busa in the chronological seat of honor, *fons et origo* of humanities computing? Or do we consider what else there might be to the confluence of Busa, Watson Sr., and the machinery (do not forget) of war and scholarship? “Yet, isn’t it true,” Busa remarks, adding yet another metaphor to our collection, “that all new ideas arise out of a *milieu* when ripe, rather than from any one individual?”

Now move to Harvard University, and fast forward to the early 1960s, when the medievalist Morton Bloomfield speculated that personification—the rhetorical device by which a poet turns subhuman entities into persons—might yield fruitful results if studied grammatically (Bloomfield 1963). His insight was that personification isn’t so much a fact as the result of a process in language by which discrete verbal factors temporarily alter the ontological state of the entity named by the noun these factors modify. (Thus when we say, “the wind sighs,” for a brief moment the action of the verb *sigh* makes the entity named by *wind* more like a person than a garden-variety breeze. But note: it does this quietly, with no obvious disturbance to the narrative—merely a suggestion that something is up.)

Little came of his speculation, and it is not difficult to see why once one takes it seriously.<sup>3</sup> When you do, attention tends to focus, as in my example, not so much on the major personification characters, which aren’t in question anyhow, rather on the short-lived, sometimes highly dubious entities that often do not survive as personifications for more than a line or two. In a long poem there may be hundreds of these, and many, many more causative factors to take into account. In every case *how much* each causative factor contributes becomes a question, since some factors are clearly stronger in that respect than others and the number of factors in each instance varies. The only reliable way of answering is by being consistent across all occurrences of each factor. It becomes obvious that even for a poem of moderate size, to produce good results rather than just idiosyncratically variable judgments at a lower level than usual, one must be able to control a large amount of data. In other words, one needs Busa’s “some type of machinery.” (As far as I know, no one reached for such machinery to implement Bloomfield’s idea until I did, a few years ago—but more about that later.) When Bloomfield was thinking about personification in the early 1960s at Harvard, about 30 minutes’ walk away Noam Chomsky was teaching at MIT, and of course much work related to computing was in full swing. The fact that Bloomfield didn’t have Busa’s inclination to follow the scent of computing, even for that short walk down Massachusetts Avenue, is irrelevant. My point, rather, is that computing was in the air, or was beginning to define the shape of the intellectual landscape, or whatever. Perhaps Bloomfield was affected without knowing it? If so, what possibly could we mean by “not knowing”? And what precisely is “it”?

#### IV. Computing’s cosmology . . . . .

By the early 1980s, when I was finishing my doctoral dissertation on Milton, microcomputers (as we called them then) were popping up everywhere, even in the humanities. I started to wonder about what a computing *of* as well as *in* the humanities might be like. Discouragingly to upstarts like myself, computing was getting *in* but it wasn’t

becoming *of*, at least not fast enough. To those in the orbit of cognitive science (which had begun about a decade earlier from the confluence of artificial intelligence, psychology, and linguistics), computing was no longer *in* the air, to be caught or not, rather it had *become* the air. In 1983, for example, the philosopher Jerry Fodor published a book, *The Modularity of Mind*, in which he took computation simply and unself-consciously *to be* the way the mind works. Nowhere in that very interesting book, and nowhere in many other discussions within and beyond cognitive science, then and now, is the equation questioned or even mentioned. To paraphrase the Canadian philosopher Charles Taylor, computation had by that point become cosmological—no longer an aspect of the world or even a worldview but how the world is unself-consciously viewed. Its success in that regard outstrips even behaviorism in its heyday (1985: 1–7).

But there were cracks in the cosmology, obvious at least to those who listened to the talk and watched the walk. What caught my attention again and again was the strategy of avoidance practiced by its believers, who would commonly use one or the other of two rhetorical moves in response to challenges coming from failures to deliver on the promise. The first was *deferral* of the date by which delivery had been promised—otherwise known as the Real Soon Now gambit. The second move was *dismissal* of whatever nasty problem or objection a skeptic might raise, or even of the entire intellectual ground on which the objector stands. Not infrequently the humanities have simply been brushed aside in that way.

What if, I thought then, one reversed the two rhetorical moves of dismissal and deferral, taking the state of computing as it now is—whenever now is, crude as it is—to be one’s epistemological instrument. (Vannevar Bush, I was pleased to discover, had compared algorithmic searching to “a stone adze in the hands of a cabinetmaker,” first in 1945, then again, “in spite of great progress,” in 1965.<sup>4</sup> That seemed about right for 1985 as well, and now, a year beyond 2005, we really should be drawing the obvious conclusion.) What if, I thought, with this admittedly crude instrument, one went after a very difficult, though tractable problem in the humanities? What kind of good would computing do? What scholarly significance would its crudities reveal?

I chose to use metalinguistic markup to classify personal references in Ovid’s *Metamorphoses* as a way of discovering structural patterns in that complex poem and as a way of putting computing to the test. These turned out to number about 60,000, that is, an average of 5 per line of poetry, and to involve very subtle distinctions as to what one called a person. In principle I could have come to the same conclusions about computing by looking at Turing’s scheme rather than at literary text on a computer, but trained as a critic, and wishing to speak intelligibly to others of the same kind, I preferred to meet the threshing machine in the cornfield rather than in the blueprint.

Two things came of my project: a hypertextual reference work, now online but never officially published, and the theoretical conclusion that for the humanities, computing’s primary gift is to problematize both methods and tools.

The realization about results led to an argument for the analytical power derived from using computers as modeling devices. Modeling, as I have come to describe it, begins when someone represents an object of study computationally, then proceeds recursively by manipulating the model, comparing the results to the original artifact and changing the model accordingly (McCarty 2005: 20–72). As it develops, the model becomes a better and better approximation, until, as is inevitable, it reveals a fundamental conceptual flaw, or raises impossible questions, and so must be scrapped in favor of a new design. Modeling is a trade-off, on the one hand requiring translation of the scholar’s idea into a radically inadequate form, on the other bestowing unprecedented ability to manipulate the data quickly on a large scale. For the humanities the process starts by privileging what we as scholars somehow know. It culminates by forcing the question of how we know what we know—to which I will return.

Since tools model methods, it follows that computing problematizes tools as well. What caused me to abandon my hypertextual reference work (called the *Analytical Onomasticon*, or “book of names”) was the realization that without software to manipulate those 60,000 tags in ways no one knows how to accommodate within current schemes, it was de facto as rigid and monocentric as any printed codex, and was therefore an unfit instrument for studying the perpetually changing expressions of Ovid’s incorrigibly

pluralizing poetic imagination. I then changed my tool of preference from encoding to relational database technology, using Bloomfield's approach, as noted earlier, to construct a modeling device for the study of personification in the *Metamorphoses*. Relational technology has given me a mature set of manipulatory tools with which to work, but at the same time it distances me from the literary text, thus greatly impeding the speed with which I can compare my analytical representation of what's happening in the text with the text itself. As Stéfan Sinclair and Julia Flanders have separately pointed out,<sup>5</sup> this distancing, imposed on us by our standard tools, is highly problematic for the study of literature. In my case, the conclusion I come to is that neither text-encoding nor relational database technology is the right kind of tool. In each case the fundamental "data model," as we might call it,<sup>6</sup> is wrong. But by this dual failure to match how literary critics read what they read, and so how I know what I know of personification in the *Metamorphoses*, computing points the way to a yet un-invented computing. And if we consider Mr. Turing's scheme, we see that this developmental metamorphosis is exactly what a computing practice of the humanities must do.

Hence my basic argument: that, by a kind of *via negativa*, failure is the key to the door from whose threshold a "great inductive leap" becomes possible. It is also the relevant phenomenological argument: that breakdown of the tool or model when we are *attending from it to the world* (as Michael Polanyi would say)<sup>7</sup> turns our relationship to it inside out and so makes possible the critical perspective its field of inquiry requires in order to mature. But, in the case of humanities computing, where does this place it in relation to the humanities? And what of the natural and mathematical sciences that computing implements and bodies forth in its influence on how we work and in the technical collaborations its use so often entails? What is our relationship to them once computing is no longer cosmological?

#### V. The humanities and the sciences. . . . .

Before attempting to answer the questions I just raised, allow me to consider a highly influential and confluent way

of talking about human work that shares the same intellectual landscape as computing.

Since the 1950s, when Alan Newell and Herbert Simon began their work in the field called "operations research" or "management science," it has been commonplace to analyze human activity by reducing it to the systematic exploration of a Cartesian "problem space."<sup>8</sup> The possible solutions in a problem space are paths within that space, and success is measurable in terms of the paths followed or the space successfully covered. As with their precursor Frederick Winslow Taylor's time and motion study of factory workers at the turn of the last century (1911), the practicality of thinking in this way depends on ignoring factors extrinsic to the scheme and on accepting the implicit proposition that entities in this space are atomic—that they are more like stars than galaxies. From an operational perspective, as problem-space exploration proceeds, the residue of misfits declines and so can be increasingly ignored, until at a certain point success may be declared.

We know that computing implements bureaucracy, of the workplace or of the mind. This is implicit in Turing's essential step, in his 1936 paper, of comparing a man solving a problem to a machine with a finite number of operations (117). Once computing enters the humanities, however, the bureaucratic misfits—those aspects of a cultural artifact that do not compute—are precisely those we want to know about. Furthermore, one or more of those misfits may, if properly understood, transform the entire problematic space. The crucial point is *not* that operationalizing, say, 99.8 percent of the problem-space turns out not to be good enough. It is that *any* residue may turn out to be, as Jerome McGann says, "the hem of a quantum garment" (2004: 201).

In terms of the metaphor, then, computing makes possible a subspace in the humanities, within which Newell's and Simon's operational strategies and computational moves obtain. Within it cultural artifacts appear *as if they were only data*, as a result of which—data being data—they may be treated *as if they were natural objects to which something like natural law applies*. The crucial matter is the *as if*, which in turn demands the bipolar ability, on the one hand to see and act on artifacts as natural objects, on the other to know them as cultural expressions. I suggested earlier that the

comparison yields epistemological insight into how we know what we know. But this operational subspace does more. As quasi-natural, it invites application of what Ian Hacking has called the “styles of scientific reasoning” (2002/1991). Just as data erases the distinction between natural and cultural, so the procedural emphasis of these “ways of being reasonable” allows them to migrate from the natural sciences into the humanities’ computational subspace.

Hacking’s styles are the discovery of Alastair Crombie, whose three-volume work, *Styles of Scientific Thinking in the European Tradition* (1994), meticulously documents the intellectual history of the natural sciences from the perspective of “how we find out, not...what we find out” (Hacking 2002: 178). Hacking brings these historical styles into the present, asking how they can be used, in words I quoted earlier, “to illumine our own predicaments.” For the humanities the first way they do this is to parallel the change in the disciplines that computing urges on us with both the historian’s and the philosopher’s common shift in attention from products to processes of reasoning. Jerome Bruner has argued that such a shift in emphasis “from the *products* of scientific and humanistic inquiry to the *processes* of inquiry themselves” has reawakened the tired old topic of the relation between the sciences and the humanities (1986: 44). More importantly, however, it has given us a new, broadly confluent way of talking about this relation.

I said earlier that the central question toward which humanities computing directs our gaze is precisely the one of epistemological process—how we know what we know. Lorraine Daston has pointed out that up to now not much attention has been paid to this question in the humanities (2004: 363). But it is important to understand why. Method has been a problem for the humanities because it tends to reduce our artifact-specific mode of work to the artifact-independent mode of the sciences (cf. Gadamer (2000/1960: 4–5). That in turn, Bruner would argue, tends to move us from emphasis on “the alternativeness of human possibility,” manifested in literature, the arts, and crafts, to a small number of universally applicable methods fit only for dealing with nature as the sciences conceive it (1986: 53). Computing’s greatest challenge to the humanities is its methodological imperative. But so long as its intellectual

subspace is understood *as if*, this challenge is no threat but an opportunity, which allows us to import and apply mature scientific methods—the gift of many centuries of hard work by brilliant people—to open up rather than close down the possibilities of things. There is, then, nothing Trojan-horsey to fear about this gift.

The second way the styles of reasoning illuminate is by directing our attention to a common emphasis on engagement with the world. *What* we find out, resulting in such things as theories or readings, is detached from its own immediate history. It stands as true or false, fruitful or fruitless, persuasive or not. But *how* we find out is all about what the finder actually *does* with available equipment. It is all about acting on the world or the artifact—hence, I think, Hacking’s preference for reasoning over thinking (which is, he says, “too much in the head”), then his regret that even reasoning gives insufficient emphasis to “the manipulative hand and the attentive eye” (2002: 180f.). Here again the least etiologically specific metaphors are likely to yield the most rewarding historiography. Here I can offer only a desultory gathering of family resemblances to suggest the extent of our constructivist leanings: the unlimited fruitfulness of Turing’s scheme; our preoccupation with constructivist theories and possible worlds; the emphasis in the natural sciences on the epistemic *Fingerspitzengefühl* of experiment;<sup>9</sup> the attention in cognitive science to what Lorenzo Magnani calls “manipulative abduction” (2002: 305); the imperative in the history of technology to learn how to read our machines (Mahoney 2003); the growth of interest in material culture (Buchli 2002), in the paradoxes of “things that talk” (Daston 2004), and the partial—but only partial—“disencumberance of meaning” as these things cross cultural boundaries (Galison 1997: 436); and, to bring this incomplete list to a close, the artisanal relation of scholar to equipment, which grows with the ongoing and much to be encouraged shift from passive *end user* to active *end maker* in the digital humanities (McCarty 2005: 15). Perhaps, as stated, little more than a cabinet of curiosities, but curiosity is a beginning, and here are many clues.

As historical rather than purely philosophical phenomena, the number and identity of the styles of reasoning cannot be prescribed. They develop in their contingent

historical contexts. So in the intellectual space I have described for the humanities, some are attested in current work, some may be in the future, others not. The modeling, statistics, and taxonomy styles, to use Hacking’s terms, are well attested. Derivation occurs, for example, in Peter Robinson’s application of cladistics, a technique from evolutionary biology, to manuscript stemmata (Robinson and O’Hara 1996). The experimental style may make sense in the humanities, particularly in its overlap with statistics and database modeling, but the details remain to be worked out. The laboratory style, involving construction of equipment to intervene in an object of study, is exemplified by Jerome McGann’s Ivanhoe Project (2005). The postulational, or style of proof, seems to have no role to play at all. For now, however, the point I wish to make is not which style is doing what, rather how their common applicability bridges the two-cultured gulf. It is a bridge built by humanists.

It is also, at least potentially, a site of interplay—a theme to which I will return. But for now, if I were, prematurely, to begin a history of humanities computing, I might start with Gottfried Wilhelm von Leibniz’s dream, “Theoreticos Empiricis felici connubio zu conjungiren,” as he said in a typical mixture of Latin and German, “to join theorists and empirics in a happy marriage” (Burke 2000: 16 f.). The metaphor is a happy one because, at least to a divorced idealist such as myself, the ideal lives side by side with experience, suitably bridged by contemporary ideas of partnership. All that aside, from Leibniz’s time, say, through Hilbert’s passion that “Wir müssen wissen—wir werden wissen!” “We must know, we shall know!”<sup>10</sup> through Gödel’s proof of truth beyond proof, to Turing’s decision-problematic machine and our own computings, runs this common though complex interplay of opposites. And though passions may run high for the one or the other, it is their creative conflict to which we must look for our history’s meaning. Opposition is true friendship.

**VI. Solidifying the ground . . . . .**

But first, as Wittgenstein said, “Zurück auf den rauhen Boden!” (*PI* 107), back to the rough ground!

My work in the modeling style, with personification, involves assigning numerical quantities to signify the personifying force of individual contributory factors. Having these quantities to work with means that I can visualize the total effect for each instance of the trope and quickly compare variations across the poem. Visualization has persuaded me that personification of the kind I study is not a binary phenomenon, as the primitive technology of capitalization denotes, but rather varies smoothly across a spectrum. Hence, when we become aware of a personification, this awareness is a matter of how we read what the poet has written, not on a fixed threshold on the wrong side of which the attempt is beneath notice. At the upper end of the spectrum, however, something very interesting happens. My attempt to account for it in the model has brought me somewhere near the middle of the bridge I was just speaking about.

Visualizing quantified personifications makes it obvious that our perception of the trope isn’t linear, and so cannot be modeled faithfully by adding up all the factors and presenting the sum. My sense of the matter is that as factors accumulate we grow accustomed to their transforming action, and so value each factor less than the one before. In other words, we saturate, just as the Russian formalists said.<sup>11</sup> (This could be tested by psychological experiment.) Saturation can be modeled mathematically using any function that dampens the accumulating sums in a reasonable way. The most interesting of the functions I know models the reception of light by the eye (McCarty 2005: 65–69). It is particularly interesting not because it produces a better dampening effect, rather because of the analogy it implies: light is to the eye as personifying stimulus is to the mind.

Before we think about the yield of this analogy, it’s important to get straight what it says. It asserts not that light is like personification and the eye like the mind, rather that the *relationships* are the same, light to eye *as* personification to mind. It is about how, not what. It is a buttoned-down metaphor.

The standard motivation for using an analogy is to probe a less well understood system by means of a better understood one. Thus, for example, Johannes Kepler reasoned that as the sun radiates light, so it must also somehow

radiate the then unknown *vis motrix*, or force causing planetary motion (Gentner 2002). But in the perceptual case, which system is the better known? As soon as one begins asking questions, it becomes clear that neurophysiological and literary processes are equally matters for research. The positivistic assumption that we passively receive and interpret a faithful image of a real object “out there” is as wrong as the information-processing notion instantiated in the model of personification I have described. Each side demands more of the other, each is (again) an improving foil to the other. Again the interplay.

Beyond that, however, is a question I will ask and simply let hang. What if the middle of the bridge is where the truth of the matter is, in this interplay between two fundamentally different ways of constructing worlds, as Bruner says (1986)—or Aristotle: on the one hand, “that which is always or for the most part” (*Met* 6.1027a20–7)—in Lorraine Daston’s words, “the oldest and, in somewhat dilute form, the most enduring” definition of the scientific (2000: 15); on the other, “not the thing that has happened, but a kind of thing that might happen” (*Poetics* 2.9.1451b). What if? I am no longer talking about a safely ensconced subspace, with trading going back and forth across a clearly marked boundary. Not *as if* now, but *what if*. What if the potential conversation between the literary critic and the neurophysiologist were the paradigm for both? What if each were each other’s metaphor?

## VII. Imaginations .....

I would hope that the anticosmologizing I have recommended, with its relating of the humanities to the sciences, is productively imaginative, but I am aware that I have not yet directly addressed the topic named in my title. I began this lecture by considering a historical imagination by which we may locate humanities computing in a much larger intellectual context, to discover, as Foucault says, not a singular identity “but a complex system of distinct and multiple elements, unable to be mastered by the powers of synthesis” (1998/1971: 386). Unable to be mastered, yes, but still interrelated, still family-resemblant. I am aware that I take

no more than baby-steps toward a historical vision of humanities computing, which to realize requires much old-fashioned *Sitzfleisch* yet to be developed, and a much broader engagement with humanity’s computings. Desire for engagement in turn leads us to the historiographical work of scholars such as Michael Mahoney for computing and technology, such as Peter Galison for the natural sciences, such as Ian Hacking for what he has apologetically called “historical ontology.” Not a synthesis, perhaps, but certainly a gathering of friends—and a great opportunity.

At the same time, although we need to learn from others, we come to the gathering by a different route, with different experiences. It is one thing to look into the sciences as a historian or a philosopher, or like Thomas Kuhn, to be now a physicist, now a philosopher, now a historian. We who talk so glibly about interdisciplinarity can all learn from his wrenching transitions across the disciplinary boundaries he traversed. Referring to the famous gestalt drawing, he describes in *The Essential Tension* seeing now the duck, now the rabbit, never a duck-rabbit (1977: 6). But it is quite another thing, calling forth very different imagery, to sail into the the archipelago of disciplines, never settling down but observing, learning, and trading. To paraphrase Isaac Deutscher’s sardonic remark as reported by Simon Schama, we have legs, not roots (1996: 29). The migrant intellectual lifestyle of humanities computing, rootless as it may be, perilous as that may be, is precisely why we have something to say, in the multicultural pidgin we have taught ourselves to speak, that no one else would ever think to say.

The thought-seed for this particular talking-back, planted some years ago by Jerome McGann, is a statement by the poet and scholar Lisa Samuels: “Beauty wedges into the artistic space a structure for continuously imagining what we do not know” (1997: 3). But what, I have been wondering, wedges into the scholarly space a structure for continuously imagining what *we* do not know? I have argued in effect that Shelley’s “creative faculty to imagine that which we know”<sup>12</sup> is not good enough, for it leaves our computings as handmaidens to old ideas and so our humanities fundamentally unchallenged. But if computing’s version of Samuels’s wedged-in space requires, as I surreptitiously suggested, an aesthetic for the landscape of confluence

from which these computings arise, then what is its language? How do we reason with it? How *do* we imagine what we don't know?

I suggested that humanities computing's subspace delimits *as if* styles of scientific reasoning which serve, *not enslave*, and that their service is realized in the scholar's mind-clearing act of denial, when his or her analytic constructs come up against the transcendent artifacts of human culture that they model. I then went on to suggest a moment ago, though much more tentatively, that the bipolar dialogue back and forth between computational model and cultural artifact may turn out to be a type of the more challenging dialogue between the adventurously interpretative humanities, such as literary criticism, and the nearby, equally adventurous natural sciences, such as neurobiology. But whether considering the former or the latter, the point is the same: that both ways of world-making are improving counter-moves to the other.

My investigations into "the imaginations of computing" began more than a year ago as a promising way of escape out from under the book I had just written. I began in good pedantic style by attempting to study "*the history of the imagination*" and soon had to abandon both definite articles. There are not merely several histories to be considered but also the larger problem of what it might mean for an idea, such as imagination, to *have* a history (Marshall 1982). The telling difficulty with my initial project, however, was the treacherous slippage (for which computing makes us prone) from the convenience of J. M. Cocking's "category of mental activity" (1991: xii–xiii) to some kind of cognitive module, reburied faculty of mind, or worse. My hopes for a history dissolved into the psychology of imagining, which in turn dissolved into the laments of some psychologists and the celebrations of others that their discipline remakes itself with every generation, if not more often, as mind reinvents mind (Jascalevich 1924; Faris 1936: 160). At the moment I rest my case with the idea I suggested a moment ago, that we imagine imagination—that it is something we do, not something we possess. Few if any definitions survive beyond the nuclear idea that imagining makes the absent present—or also, Robert Asen points out, the present absent (2002: 355). Perhaps Wallace Stevens's

"the power of the mind over the possibilities of things" captures what we mean by the word now (1951/1949: 136). We imagine imagination afresh, my research suggests, when the space we're in proves too small, or when some residue we have not swept under the procedural carpet trips us up, and so wakes us up. We imagine imagination by means of a foil, which is precisely why the strength of the foil and the intellectual vigor of he or she who wields it is so vital, why its opposition is truer friendship than we could hope to find anywhere else.

I end with the Irish poet Seamus Heaney's poem *Field of Vision*.

I remember this woman who sat for years  
In a wheelchair, looking straight ahead  
Out the window at sycamore trees unleafing  
And leafing at the far end of the lane.

Straight out past the TV in the corner,  
The stunted, agitated hawthorn bush,  
The same small calves with their backs to wind and rain,  
The same acre of ragwort, the same mountain.

She was steadfast as the big window itself.  
Her brow was clear as the chrome bits of the chair.  
She never lamented once and she never  
Carried a spare ounce of emotional weight.

Face to face with her was an education  
Of the sort you got across a well-braced gate—  
One of those lean, clean, iron, roadside ones  
Between two whitewashed pillars, where you could see

Deeper into the country than you expected  
And discovered that the field behind the hedge  
Grew more distinctly strange as you kept standing  
Focused and drawn in by what barred the way.

## Notes

- <sup>1</sup> For the history of recent science, see McCarty 2004.
- <sup>2</sup> Hacking 2002: 182; Roth 1981.
- <sup>3</sup> For the relevant literature, see Bloomfield 1963; McCarty 1993: 136 n. 33.
- <sup>4</sup> Bush 1945: 105 and 1967: 92, both reprinted in Nyce and Kahn, eds. 1991 (pp. 99 and 209, respectively). As Bush notes at the beginning of the 1967 article, “Memex Revisited,” it was written in 1965 but not published until 1967.
- <sup>5</sup> Sinclair 2003: 180; Flanders 2005: 54.
- <sup>6</sup> I am being tentative here because “data model” is a technical term in relational database theory and was first defined by E. F. Codd (1980). One might better say, “a formal way of working with computers.”
- <sup>7</sup> Polanyi 1983/1966.
- <sup>8</sup> Newell and Simon 1972; see Simon and Newell 1958 and Popper 1999/1991; contrast Polanyi 1957.
- <sup>9</sup> *Fingerspitzengefühl* is variously translated as “firsthand experience,” “gut-instinct,” or “intuitive feeling,” but is best rendered literally, “feeling on the fingertips,” which captures both the tacit quality of knowledge obtained physically and the body-part directly involved in much experimental work.
- <sup>10</sup> The inscription on Hilbert’s tombstone, from his radio address, Hilbert 1930.
- <sup>11</sup> Shlovsky 1965/1917; Jakobson 1987/1960; cf. Bruner 1983.
- <sup>12</sup> *Defense of Poetry* (1821), Part I.38.270.

## Works Cited

- Asen, Robert. 2002. “Imagining in the Public Sphere.” *Philosophy and Rhetoric* 35.4: 345–67.
- Atran, Scott. 1996/1990. *Cognitive Foundations of Natural History: Towards an Anthropology of Science*. Cambridge: Cambridge University Press.
- Bloomfield, Morton W. 1963. “Grammatical Approach to Personification Allegory.” *Modern Philology* 60.3: 161–71.
- Bruner, Jerome. 1986. “Possible Castles.” In *Actual Minds, Possible Worlds*. 44–54. Cambridge, MA: Harvard University Press.
- Buchli, Victor, ed. 2002. *The Material Culture Reader*. Oxford: Berg.
- Burke, Peter. 2000. *A Social History of Knowledge from Gutenberg to Diderot*. Vonhoff Lectures, University of Groningen. Cambridge: Polity Press.
- Busa, Roberto. 1980. “The Annals of Humanities Computing: The *Index Thomisticus*.” *Computers and the Humanities* 14: 83–90.
- Bush, Vannevar. 1945. “As We May Think.” *Atlantic Monthly*, July: 101–8.
- \_\_\_\_\_. 1967. “Memex Revisited.” In *Science Is Not Enough*. 75–101. New York: William Morrow.
- Cocking, J. M. 1991. *Imagination: A Study in the History of Ideas*. Ed. Penelope Murray. London: Routledge.
- Codd, E. F. 1980. “Data Models in Database Management.” *ACM SIGMOD Record: Proceedings of the 1980 Workshop on Data Abstraction, Databases and Conceptual Modeling* 11.2: 112–14.
- Crombie, A. C. 1994. *Styles of Scientific Thinking in the European Tradition: The History of Argument and Explanation Especially in the Mathematical and Biomedical Sciences and Arts*. 3 vols. London: Duckworth.
- Daston, Lorraine. 2000. “Preternatural Philosophy.” In *Biographies of Scientific Objects*. Ed. Lorraine Daston. 15–41. Chicago: University of Chicago Press.
- \_\_\_\_\_. 2004. “Whither CI?” *Critical Inquiry* 30.2: 361–4.
- \_\_\_\_\_, ed. 2004. *Things That Talk: Object Lessons from Art and Science*. New York: Zone Books.
- Ewald, William, ed. 1996. *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*. Vol. 2. 2 vols. Oxford Science Publications. Oxford: Clarendon Press.
- Flanders, Julia. 2005. “Detailism, Digital Texts, and the Problem of Pedantry.” *Text Technology* 14.2: 41–70.
- Fodor, Jerry A. 1983. *Modularity of Mind*. Cambridge, MA: MIT Press.
- Frye, Northrop. 1957. *Anatomy of Criticism: Four Essays*. Princeton: Princeton University Press.
- Gadamer, H.-G. 2000/1960. *Truth and Method*. 2nd ed. Trans. Joel Weinsheimer and Donald G. Marshall. New York: Continuum.
- Galison, Peter. 1997. *Image and Logic: A Material Culture of Microphysics*. Chicago: University of Chicago Press.
- Gandy, Robin. 1995/1994. “The Confluence of Ideas in 1936.” In *The Universal Turing Machine—A Half-Century Survey*. Ed. Rolf Herken. 51–102. Computerkultur II. 2nd ed. Wien: Springer Verlag.
- Geertz, Clifford. 2000/1983. “Anti Anti-Relativism.” In *Available Light: Anthropological Reflections on Philosophical Topics*. 42–67. Princeton: Princeton University Press.
- Gentner, Dedre. 2002. “Analogy in Scientific Discovery: The Case of Johannes Kepler.” In Magnani and Nersessian 2002: 21–39.
- Gödel, Kurt. 1995/ca. 1961. “The Modern Development of the Foundations of Mathematics in the Light of Philosophy.” In *Unpublished Essays and Lectures*. Ed. Solomon Feferman, John W. Dawson, Jr., Warren Goldfarb, Charles Parsons, and Robert M. Solovay. 364–87. Vol. 3 of *Collected Works*. 3 vols. Oxford: Oxford University Press.
- Hacking, Ian. 2002/1991. “‘Style’ for Historians and Philosophers.” In *Historical Ontology*. 178–99. Cambridge, MA: Harvard University Press.
- Hilbert, David. 1902. “Mathematical Problems.” *Bulletin of the American Mathematical Society* 37.4: 407–36.
- \_\_\_\_\_. 1930. “Naturerkennen und Logik.” In Viktor Vinnikov, “We Shall Know: Hilbert’s Apology.” *Mathematical Intelligencer* 21 (1999): 42–46.
- Jakobson, Roman. 1987/1960. “Linguistics and Poetics.” In *Language in Literature*. Ed. Krystyna Pomorska and Stephen Rudy. 62–94. Cambridge, MA: Harvard University Press.
- Jascalevich, Alexander A. 1924. “The Idea of Continuity in the History of Psychology I.” *Journal of Philosophy* 21.24: 645–63.
- Kearney, Richard. 1988. *The Wake of Imagination: Ideas of Creativity in Western Culture*. London: Hutchinson.
- Kuhn, Thomas S. 1977. *The Essential Tension: Selected Studies in Scientific Tradition and Change*. Chicago: University of Chicago Press.
- Magnani, Lorenzo. 2002. “Epistemic Mediators and Model-Based Discovery in Science.” In Magnani and Nersessian 2002: 305–29.

- \_\_\_\_\_, and Nancy J. Nersessian, eds. 2002. *Model-Based Reasoning: Science, Technology, Values*. New York: Kluwer Academic / Plenum Publishers.
- Mahoney, Michael S. 2003. "Reading a Machine." [www.princeton.edu/~hos/h398/readmach/modelt.html](http://www.princeton.edu/~hos/h398/readmach/modelt.html) (22 October 2006).
- Marshall, Donald G. 1982. "Ideas and History: The Case of 'Imagination.'" *Boundary 2* 10.3: 343–59.
- McCarty, Willard. 1993. "Encoding Persons and Places in the *Metamorphoses* of Ovid: 1. Engineering the Text." *Texte: Revue de critique et de théorie littéraire* 13/14: 121–72.
- \_\_\_\_\_. 2004. "As It Almost Was: Historiography of Recent Things." *Literary and Linguistic Computing* 19.2: 161–80.
- \_\_\_\_\_. 2005. *Humanities Computing*. Basingstoke: Palgrave.
- McGann, Jerome. 2004. "Marking Texts of Many Dimensions." In *A Companion to Digital Humanities*. Ed. Susan Schreibman, Ray Siemens, and John Unsworth. 198–217. Oxford: Blackwell.
- \_\_\_\_\_. 2005. "Ivanhoe: An Interface for Interpretation." *Literary Magazine* 1.1. [www.litencyc.com/theliterarymagazine/ivanhoeinterface.php](http://www.litencyc.com/theliterarymagazine/ivanhoeinterface.php) (22 October 2006).
- Newell, Allen, and Herbert A. Simon. 1972. *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Nyce, James M., and Paul Kahn, eds. 1991. *From Memex to Hypertext: Vannevar Bush and the Mind's Machine*. Boston: Academic Press.
- Polanyi, Michael. 1957. "Problem Solving." *British Journal for the Philosophy of Science* 8.30: 89–103.
- \_\_\_\_\_. 1983/1966. *The Tacit Dimension*. Gloucester, MA: Peter Smith.
- Popper, Karl. 1999/1991. "All Life Is Problem Solving." In *All Life Is Problem Solving*. Trans. Patrick Camiller. 99–104. London: Routledge.
- Prown, Jules David. 1988. "Mind in Matter: An Introduction to Material Culture, Theory and Method." In *Material Life in America, 1600–1800*. Ed. Robert Blair St. George. Boston: Northeastern University Press.
- Robinson, Peter M. W., and Robert J. O'Hara. 1996. "Cladistic Analysis of an Old Norse Manuscript Tradition." *Research in Humanities Computing* 4. *Selected Papers from the ALLC/ACH Conference, Christ Church, Oxford, April 1992*. 115–37. Oxford: Clarendon Press.
- Rorty, Richard. 1980. *Philosophy and the Mirror of Nature*. Princeton: Princeton University Press.
- Roth, Michael S. 1981. "Foucault's 'History of the Present.'" *History and Theory* 20.1 (1981): 32–46.
- Samuels, Lisa. 1977. Introduction to special issue, *Poetry and the Problem of Beauty*. *Modern Language Studies* 27.2: 3–6.
- Schama, Simon. 1996. *Landscape and Memory*. New York: Vintage Books.
- Shelley, Percy Bysshe. 1821. *Defense of Poetry*.
- Shlovsky, Viktor. 1965/1917. "Art as Technique." In *Russian Formalist Criticism: Four Essays*. Ed. and trans. Lee T. Lemon and Marion J. Reis. 3–24. Lincoln: University of Nebraska Press.
- Simon, Herbert A., and Allen Newell. 1958. "Heuristic Problem Solving: The Next Advance in Operations Research." *Operations Research* 6: 1–10.
- Sinclair, Stéfan. 2003. "Computer-Assisted Reading: Reconceiving Text Analysis." *Literary and Linguistic Computing* 18.2: 175–84.
- Stevens, Wallace. 1951/1949. "Imagination as Value." In *The Necessary Angel: Essays on Reality and the Imagination*. 131–56. New York: Vintage Books.
- Taylor, Charles. 1985. *Philosophy and the Human Sciences*. Vol. 2 of *Philosophical Papers*. 2 vols. Cambridge: Cambridge University Press.
- Taylor, Frederick W. 1911. *The Principles of Scientific Management*. New York: Harper and Brothers.
- Turing, Alan. 1936. "On Computable Numbers, with an Application to the Entscheidungsproblem." *Proceedings of the London Mathematical Society* 42: 230–65.
- Winner, Langdon. 1997/1986. "Technologies as Forms of Life." In *Technology and Values*. Ed. Laura Westra and Kristin Shrader-Frechette. 55–69. Oxford: Rowman and Littlefield.
- Wittgenstein, Ludwig. 2001/1953. *Philosophical Investigations*. Trans. G. E. M. Anscombe. Oxford: Blackwell.



The National Humanities Center is the only major independent American institute for advanced study in all fields of the humanities. Privately incorporated and governed by a distinguished board of trustees from academic, professional, and public life, the Center was planned under the auspices of the American Academy of Arts and Sciences and began operation in 1978. It provides a national focus for the best work in the liberal arts, drawing attention to the enduring value of ancient and modern history, language and literature, ethical and moral reflection, artistic and cultural traditions, and critical thought in every area of humanistic investigation. By encouraging excellence in scholarship, the Center seeks to ensure the continuing strength of the liberal arts and to affirm the importance of the humanities in American life.

#### Trustees of the Center

Carl H. Pforzheimer III  
*Chairman*  
Geoffrey G. Harpham  
*President*  
Peter A. Benoiel  
*Vice Chairman*  
Steven Marcus  
*Vice Chairman*  
Patricia Meyer Spacks  
*Vice Chairman*  
John F. Adams  
*Secretary*  
Merril M. Halpern  
*Treasurer*

James H. Averill  
Enriqueta C. Bond  
Erskine Bowles  
Alan Brinkley  
Richard H. Brodhead  
Cees J. de Bruin  
Frances Ferguson  
Frances Daly Fergusson  
Catherine Gallagher  
Michael Ann Holly  
Kimberly J. Jenkins

William Johnson  
William C. Jordan  
Stanhope A. Kelly  
Philip S. Khoury  
Thomas Laqueur  
Assad Meymandi, M.D.,  
Ph.D.  
Abby S. Milstein  
William M. Moore, Jr.  
Patricia R. Morton  
Robert C. Post  
Hunter Rawlings  
Bruce Redford  
Lawrence R. Ricciardi  
Sally Dalton Robinson  
Thomas Scherer  
Lanty L. Smith  
Herbert S. Winokur, Jr.  
Pauline R. Yu

#### Trustees Emeriti

Meyer H. Abrams  
John Birkelund  
William J. Bouwsma\* †  
Frederick Burkhardt †  
Caroline W. Bynum  
W. Robert Connor  
Andrew H. Delbanco\*  
Strachan Donnelley  
Blair Effron  
Jean Bethke Elshtain\*  
Nancy Faircloth  
John Hope Franklin\*  
William Friday  
Robert Goheen  
William Golden †  
Caryl Haskins †  
Gertrude Himmelfarb  
Robert Hollander  
Dan Lacy †  
William Leuchtenberg\*  
Edward Levi †  
Martin Marty  
Claude McKinney  
John G. Medlin, Jr.  
J. Irwin Miller †

Francis Oakley\*  
John Oates †  
Colin Palmer  
Benno Schmidt  
Anne Firor Scott\*  
John Searle  
Isaac Shapiro  
John Voss  
Stephen Weiss †

#### Founders

Meyer H. Abrams  
Morton Bloomfield †  
Frederick Burkhardt †  
Charles Frankel †  
*First Director, 1977–79*  
Robert F. Goheen †  
Steven Marcus  
Henry Nash Smith †  
Gregory Vlastos †  
John Voss †

\* Fellow † Deceased



National Humanities Center  
7 T. W. Alexander Drive  
Research Triangle Park, NC 27709  
[nationalhumanitiescenter.org](http://nationalhumanitiescenter.org)  
(919) 549-0661