

BRAVE NEW WORLD:
A LOOK AT *18THCONNECT*

LAURA MANDELL

The rapidly changing scholarly digital universe requires both new tools and new practices to meet the challenges it presents. Because the skill sets required to design the digital universe vary, scholars, librarians, computer scientists, and even administrators must work together to create platforms for locating, searching, and accessing the data now stored in electronic media. If scholars fail to join librarians in this task, many valuable digital materials will effectively vanish, their traces washed away in the data deluge. Similarly, institutions must adjust their practices to meet the realities of the new digital world, especially with regard to approaches to tenure and promotion. These institutional practices require a vetting process equivalent to juried publication for assessing the quality of novel forms of scholarship such as databases, electronic editions, software, finding aids, or digital tools. What is needed is a collaborative effort that permits young scholars with savvy ideas for digital projects to develop those projects, confident that their work will receive the peer review necessary for keeping their jobs. Finally, there is a larger systemic issue that Jerome McGann notes: our cultural heritage is going to be digitized, and scholars need to be at the table when people are deciding when, where, and how.¹

To address these needs, I founded and currently co-direct *18thConnect* (<http://www.18thConnect.org>), an exciting new community, electronic platform, and research portal currently under construction and supported by Miami University of Ohio, the University of Virginia, an NEH grant from the Institute for Computing in the Humanities, Arts, and Social Sciences (I-CHASS) at the University of Illinois, and a grant from the Andrew W. Mellon Foundation awarded to Miami University, my home institution. Brad Pasanek of the University of Virginia co-directs *18thConnect*, which takes its

inspiration from *NINES* (*Nineteenth-Century Scholarship Online*) founded by Jerome McGann and directed by Andrew Stauffer, also at the University of Virginia. Like *NINES*, *18thConnect* will offer scholars access to electronic resources by providing an aggregated integration of those resources. In other words, *18thConnect* will serve as a single platform through which multiple resources can be searched. Freely available databases, such as the *English Short-Title Catalogue* (*ESTC*), the *Old Bailey Online*, and the 2,180 *ECCO* texts that have been transcribed by Michigan's Text Creation Partnership, can be consulted with a simple click on a link returned in a search. While proprietary resources can be fully accessed only by those whose institutions subscribe to them, all searches will return bibliographic data into the *18thConnect* portal. *18thConnect* has also secured agreements with commercial vendors to include the *Cambridge Bibliography of English Literature*, Gale-Cengage's *Eighteenth-Century Collections Online* (*ECCO*), and Adam Matthew's *Eighteenth-Century Journals Portal* (*E-C Journals*).² The editorial board and steering committee, moreover, will be actively soliciting submissions from additional digital resources in the fields of art history, literature, history, and philosophy.

18thConnect peer-reviews each database and electronic scholarly edition in its collection. In fact, one of *18thConnect*'s tasks will be to serve as a vetting agent, helping to legitimize worthy electronic scholarship for promotion and tenure committees. The editorial board members of *18thConnect* are as illustrious as those of any major press. When *18thConnect* accepts an electronic scholarly edition, it will provide letters of acceptance designed to help interpret an accepted project's value for promotion and tenure committee members for whom such work may be new.³ *18thConnect* follows in the tradition established by Jerome McGann's purpose in establishing *NINES* for nineteenth-century scholarship (<http://www.nines.org>). Both projects ensure that high-quality scholarship is distinguished from the mass of materials available online; both ensure that promising young scholars will receive the credit they need and deserve for high quality work.⁴ *NINES* has received an NEH Summer Institute Grant to conduct two institutes on this very topic: representatives from the ACLS and MLA participating in this *NINES* Summer Institute will work together with chairs, administrators, and digital humanists in order to come up with criteria for both judging work and communicating its value to promotion and tenure committees.

It should be emphasized that *18thConnect* serves only as an *aggregator* for online projects; the databases, editions, and electronic scholarship to which it offers access remain in the hands of their creators and will not be owned, controlled, or operated by *18thConnect*. Open-access sites thus retain their creative freedom and can be submitted for peer-review while still undergoing major development. In fact, it is ideal for creators or editors to submit their

work in its developmental stage: *18thConnect* and *NINES* together offer instruction via summer workshops and institutes on how to create library-quality, state-of-the-art digital resources, projects worthy of becoming, for instance, MLA Electronic Scholarly Editions. Though *18thConnect* only points to rather than ingests digital scholarship, proprietary collections and journals such as *ECCO*, *E-C Journals*, and *JSTOR* are searchable through *18thConnect*. Full text can be searched when it has been made available by the owners or creators of the resource. Users are returned text snippets, and then can click on the title of the work to see the full article, digital facsimile, electronic edition, or data strand. As mentioned above, freely available texts are completely accessible through *18thConnect*, whereas texts in proprietary collections such as *ECCO* or the *E-C Journals* portal are accessible to users only if their home institutions subscribe. Users with access to *ECCO* at work or home but using a proxy server can immediately access an *ECCO* text by searching *18thConnect*. Once *18thConnect* is fully functional, a link to that primary text will be returned, accompanied by links to all the reliable scholarly information available on the web related to the search term or title—from scholarly articles to versions, editions, images, and the like.

What of those whose institutions do not subscribe to *ECCO*, the major eighteenth-century resource for primary materials, containing literally millions of page images of texts in the fields of language, literature, law, social science, medicine, etc.? *18thConnect* has worked out what may prove to be a historic agreement with Gale–Cengage that will provide search access and more to users whose institutions cannot afford to subscribe.

Gale's *ECCO* contains page images for over 182,000 texts, some of them multi-volume texts as lengthy as *Clarissa*. Creating such a set of images has taken decades of work, but some of the page images are not sufficiently readable to be transformed into typed texts by computer programs designed for this work. Earlier OCR (Optical Character Recognition) software was unequipped to handle the variability of eighteenth-century typography, often misreading words and thereby compromising text searches. *ECCO*'s digitized images, moreover, were made from microfilmed copies of the original titles, and this also sometimes compromises the legibility of the scanned page. To address this problem, new, more precise OCR software is needed. Grants from the Mellon Foundation and NEH (through the National Center for Supercomputer Applications [NCSA] and I-CHASS) are funding Miami University's development of a new, open-source software program for mechanically transforming images into typed texts. *18thConnect* will re-run *ECCO* page images through this new program in order to generate cleaner text than Gale has been able to produce so far. Next, *18thConnect* will provide a window for

users—anyone who wishes to register with an e-mail address—to correct the typing of these texts.

The new OCR software we will use, Gamera, is an open-source program originally developed by Professor Ichiro Fujinaga of McGill University and published by Johns Hopkins University (<http://gamera.informatik.hsr.de/>). Because Gamera was originally created for recognizing musical characters, it is less dependent than other OCR software on recognizing characters only if they occur on the same line as others. This feature is valuable for scanning texts produced before 1820 because the characters in those texts are often not evenly aligned along a baseline, the result of the punch not being situated in the matrix with mathematical precision when the type was made. We have already been able to train Gamera to distinguish between the long *s* and the lowercase *f*, something that was previously possible only through dictionary look-up. There are, however, some things Gamera may not do as well as Gale's OCR, so we are further developing automated correction, and the centerpiece of our process is a crowd-sourced correction tool. It is time, Martin Mueller has said, for scholars to wash their own dishes: the more scholars help us correct texts, the better scholarly searches will be, and ultimately—because everything will be sent back to Gale—the better the archive will be for future scholars.⁵ That these texts be correctly typed is crucial for searching and data-mining; only by providing texts that can be accurately read by machines will we make them locatable, usable, and comprehensible to future generations.

Contributing to the future robustness and integrity of the archive, users of *18thConnect* can search for documents in the *ECCO* collection and (1) correct errors found in the snippets, or (2) register in order to see and read portions of a text in exchange for correcting it. If someone has registered as a user at *18thConnect*, which requires only a username and e-mail address, that person can save documents that he or she would like to correct into a personal account on the "My18" page. Additionally, if a user decides to correct a whole text, once the corrections are completed through the online correction tool accessible through the "My18" page, *18thConnect* will immediately give the correctly typed version of that document to the person who corrected it *to use as he or she likes*, and we will do so in several forms: both plain text and text encoded in TEI, which the MLA requires for electronic scholarly editions.⁶ We will also provide guidance so that users can create a library-quality digital edition and use all the newest tools on their documents such as JuXta and the Versioning Machine, which enable one to compare various editions, as well as exciting new visualization tools such as TAPor, Voyeur: Reveal Your Texts, and TokenX.⁷ Many of these tools will be made accessible through *18thConnect* in 2011. If sufficiently researched and annotated, editions built with our help can be submitted to *18thConnect* for peer review. Furthermore, library-quality

scholarly editions are eligible to become MLA Electronic Scholarly Editions. Positively reviewed editions are first accepted into the *18thConnect* online finding aid. If a scholar's edition has been accepted (positively peer-reviewed), Gale–Cengage may choose to publish the edition along with the page images as a print-on-demand edition, or other print-on-demand publishers can be enlisted to print the transcribed, annotated edition.⁸ Thus, in exchange for correcting texts and improving the archive, users will have produced a scholarly resource that will count toward tenure and that is both digital edition and printed book.

18thConnect will provide coded documents, instructions, and support so that scholarship produced by users—ideally professors of eighteenth-century literature, culture, history, and art—is of the highest digital quality. The TEI coding that we provide ensures that these books can be produced in any number of forms—not just web pages but E-books, for instance, and whatever forms become necessary as new technologies emerge. In other words, one's scholarly work will be preserved for future use in a way that book publishing no longer guarantees. Though print copies will continue to exist, the standard means for discovering and searching scholarly work will be electronic; only by making one's work fully searchable by word will it rise to the surface of searches.

Let me give one example, situating the problem in recent debates about the scholarly uses of digital resources. John Guillory has taken issue with N. Katherine Hayles's praise of multitasking forms of attention, insisting on the value of the sustained attention required for close reading.⁹ Multi-tasking is associated with the digital, as is the "distant reading" proposed by Franco Moretti in his *Graphs, Maps, and Trees*, as well as elsewhere in his arguments with Katie Trumpener.¹⁰ The activity of distant reading involves looking at visual representations of searching and data-mining many texts, and it can be supplemented by close reading in and around texts discovered to be most interesting. But as is plain in reading an explicit account of the reading practices of scholars at our moment, neither close reading nor distant reading alone fully captures what we do.¹¹ Scholars always have, and ideally always will, act as "filters" of information based on minds developed through discipline and intensive study that cannot be duplicated by machines. While that will not change, what will change in the new digital universe is the advent of new and powerful search functions that will become indispensable, both for the cognitive filtering traditional scholarship demands and for the proper ordering and cataloguing of new digital archives, making findable the information they hold.

A search that I conducted in *18thConnect* further illustrates this point. I have been interested in historically changing notions of factuality and began

searching for a phrase I had encountered in my reading that I thought might be a forerunner to the legal notion of “circumstantial evidence.” My goal was to investigate whether the evidence of circumstance was as devalued during the eighteenth century as it is now in the phrase, “that’s just circumstantial” — i.e., won’t hold up in a court of law. The phrase that might be a forerunner was “circumstantial information”; as eighteenth-century scholars know, “to give an information on” someone meant to testify against that person, and usually to legal authorities of some sort. My search in Gale’s *ECCO* returned thirty-six results:

The screenshot displays the Gale Eighteenth Century Collections Online (ECCO) search results page. At the top, there is a navigation bar with links for 'CHANGE DATABASES', 'HELP', 'LOGIN', 'MIAMI UNIVERSITY MAIN CAMPUS OXFORD', 'RETURN TO LIBRARY', and 'LOG OUT'. Below this is the 'Eighteenth Century Collections Online' header with a search bar and a 'Search' button. A secondary navigation bar includes 'HOME', 'ADVANCED SEARCH', 'BROWSE AUTHORS', 'BROWSE WORKS', 'RESEARCH TOOLS', and 'SEARCH HISTORY'. The main content area shows 'MARKED ITEMS(2)' and 'ADVANCED SEARCH > RESULTS'. The search results are displayed in a list format, with each entry including a checkbox, a title, author, and details. The search criteria are shown as 'Results for Advanced Search (Entire Document=(“circumstantial information”))LIMITS:(Literature and Language)'. The search results are sorted by 'Author' and show 1-6 of 6 results. The first two results are by Alexander Campbell, and the last two are by Frederick II, King of Prussia.

Search within these results	Mark All	Sort By: Author
GO	<input type="checkbox"/>	Showing 1-6 of 6 1
Narrow results by subject area:		
Literature and Language (6)		
<input type="checkbox"/>	TITLE: An introduction to the history of poetry in Scotland, from the beginning of the thirteenth century down to the present time; together with a...	AUTHOR: Campbell, Alexander. DETAIL: Edinburgh, 1798. 628pp. Literature and Language Full Citation: Table of Contents List of Illustrations Keyword in Context
<input type="checkbox"/>	TITLE: An introduction to the history of poetry in Scotland, from the beginning of the thirteenth century down to the present time; together with a...	AUTHOR: Campbell, Alexander. DETAIL: Edinburgh, M.DCC.XCVIII. [1798][1799]. 630pp. Literature and Language Full Citation: Table of Contents List of Illustrations Keyword in Context
<input type="checkbox"/>	TITLE: Correspondence, Letters between Frederic II, the Prince of Prussia, and General Fouquet. Miscellanies. Translated from the French by Thomas Holcroft.	DETAIL: London, M.DCC.LXXXIX. [1789]. 520pp. Literature and Language Full Citation: Table of Contents Keyword in Context
<input type="checkbox"/>	TITLE: Political, philosophical, and satirical miscellanies. Translated from the French by Thomas Holcroft.	AUTHOR: Frederick II, King of Prussia. DETAIL: London, M.DCC.LXXXIX. [1789]. 385pp. Literature and Language Full Citation: Table of Contents Keyword in Context

Figure 5: ECCO.

As I scrolled through the six pages of returns, I began to classify them—not in the way that Gale does, based on its generalized textual categories. Instead, I classified by content, separating those entries having to do with legal or historical testimony when some account’s veracity was in doubt (20), from those discussing evidence for the existence of God (4), from those having to do with Frederick the Great (3), from works either translated or written by Thomas Holcroft, the lawyer (3). I did not see any further pattern in the materials. There were memoirs (2), and two editions of a history of “Public Characters,” again suggesting that justification in some kind of dispute among public figures was being recounted. Finally, there were some interesting individual instances, such as Thomas Clarkson’s treatise on slavery, Arthur

Young's *Rural Oeconomy*, and one complete anomaly, Alexander Campbell's *Introduction to the History of Poetry in Scotland*—some poet must have been involved in a dispute, I reasoned.

My next question was technical: would any additional search results from ECCO be returned if the searchable text (the automatically generated text running behind the page images, and making them searchable) were more correctly typed? I went into 18thConnect to search ECCO, limiting the search to “full-text only.” The full 182,000 texts, transcribed a little more accurately by our Gamera OCR engine, might have given us a few more returns as well, but I tried it simply on the 2,180 hand-typed texts currently available in 18thConnect, or just a little over 1 percent of the full Gale ECCO collection. I got one return, but, significantly, it differed from the ECCO returns:

The screenshot shows the 18thConnect search interface. The search query is "ECCO" and the search term is "circumstantial information". The search results show one entry: "Literary memoirs of living authors of Great Britain, arranged according to an alphabetical catalogue of their Names; and including a list of their works, with occasional opinions upon their literary character. In two volumes. ..." by Rivers, David. The page also shows a sidebar with "Currently Searching..." and "Limit Results to..." options.

Resource	# of Objects
18thConnect	81,855
NINES	331,816

Resource	# of Objects
Projects (Peer Reviewed)	1,386
Other Digital Collections	80,469

Genre	# of Objects
Book History	12
Citation	2,180
Collection	117
Criticism	18
Drama	65
Education	6
Education	6

Figure 6: 18thConnect.

(Please see the note about this search.)¹² The return that came from using more correctly typed text is the *Literary Memoirs of Living Authors of Great Britain* by David Rivers, published in Calcutta, which intones in its preface, “This work contains much excellent and circumstantial information related in a strain of impartiality” I immediately searched 18thConnect again for “circumstantial,” “information,” and “literary” to see whether there was some unexpected

connection between this phrase and literary biographies. I discovered that in “An Essay on the Study of Literature” Edward Gibbon praises “A circumstantial knowlege [*sic*] of [the] situation and manners” as necessary for understanding texts written in “another clime and born in another age.” I checked the whole of *18thConnect* for “circumstantial evidence,” and received returns from the *Old Bailey Online*, in which the phrase has its modern sense. In these trial accounts, one can even find “circumstantial evidence” deemed too weak, allowing a suspect to be released. But as is immediately visible in the snippet views of texts returned in *18thConnect*, there are distinctly literary uses in which this kind of “information” or evidence is accounted valuable. After this search, I went back to the results from Gale’s *ECCO* and saw another pattern that gathers together texts previously categorized as anomalous, memoir, and state biography: in tracing the occurrence of the phrase in those texts, I discovered that “circumstantial information” is presented as crucial to literary biography as well as literary history. This account provides just one example of the kinds of analysis that will be possible once good, reliable, and thus searchable texts are made possible through *18thConnect*.

As an online finding aid, and a crowd-sourced correction tool, *18thConnect* forges a beneficial connection between those who care deeply about future scholarship and the librarians and vendors who are creating the digital archive where, indeed, all that future scholarship will take place.

But what about now?

If you go to *18thConnect* today, go to “advanced search,” click on “*ECCO*” under “collections,” and then, at the bottom of the search selection menu found on the right, click on “full text only,” anyone—registered or not—can perform a full-text search by word of 2,180 texts in the *ECCO* collection. These texts have been produced by the Text Creation Partnership at the University of Michigan. They have been triple-keyed, which means typed three times and compared, a way of ensuring accuracy.¹³

In addition to conducting full-text searches on very accurate texts in *18thConnect*—more of which will be added, of course, as we enlist our users in correcting them—one can also perform two other tasks. Researchers using this online finding aid can write articles within the resource itself by using the *NINES/18thConnect* “exhibit builder.” Users can search, collect, tag, and take notes on items in the database, which currently contains 670,000 items and is growing. Once logged in and working from the “My 18” page, one can write an article that incorporates these items, either as bibliographic citations or as images. One might include a Caribbean map, for instance, in an article about circum-Atlantic circulation. If the author clicks on “publish,” this article can be submitted for peer review. Non-peer-reviewed publication is possible as well, so that professors can use exhibits as syllabi or articles for classes, and students

can be asked to build exhibits.¹⁴ This exhibit-building space is a composition environment that functions like a blog entry, and it is as easy to use as Microsoft Word. Second, *18thConnect* is a research environment that is both private and communal. Tagging, note-taking, and saving searches are possible on *18thConnect*, and tags are shared with the general community. One can even subscribe to a tag—"Melancholy," for example—helpful to someone doing research on that topic. Receiving regular feeds about how other scholars are employing a specific tag promises to engage researchers in current, broader conversations about a given topic.

We have plans for future functionalities, including a window that will allow for fully analyzing the primary texts that have been typed and corrected, as well as any secondary materials. This window is called "Voyeur: Reveal Your Data" (see note 7), a program developed by Canadian researchers, Geoffrey Rockwell and Stéfán Sinclair. "Voyeur" will allow writers to include "live data pours" in their digital essays, and *18thConnect* has already enlisted top-notch scholars to perform data mining using this tool and to make movies or screencasts that will be available on *18thConnect*. These screencasts will instruct scholars and students in how to do the same. Mellon is funding Miami University to undertake this initiative and build the crowd-sourced correction tool.

What kinds of projects will be submitted to *18thConnect* once it is up and running? The 2011 MLA panel "Digital Lives," sponsored by the division of eighteenth-century comparative literature and chaired by Ruth Mack, illustrated two of the major digital projects of the decade, both of which won NEH's "Digging into Data Challenge Grant." Significantly, both are in fact eighteenth-century projects: *Mapping the Enlightenment through Letters* (Stanford University) and *London Lives* (Sheffield University).¹⁵ The panel offered detailed instructions about how to use these resources.

More than anything, *18thConnect* aspires to be a community of scholars working together to make our research and publication environments all that they can be, all that we want them to be. Currently, many valuable eighteenth-century digital projects exist, but they exist in their isolated virtual space in an unordered digital universe. By providing a single portal through which these projects can be accessed, *18thConnect* will facilitate dialogue among the scholars behind such projects and make the projects themselves both searchable and minable at the same time, in one place, together. I hope sincerely that you will join, explore, critique, give us feedback, try to break our tools (so that we can make them better), find missing information, publish, surf, and enjoy: <http://www.18thConnect.org>.

NOTES

I would like to thank Anna Battigelli and Eleanor Shevlin: without their hard work, this essay would not have been written.

1. Jerome McGann, "Remarks," Roundtable on Electronic Editions and Archives of Poetry, Modern Language Association, 29 Dec. 2008.
2. *The Eighteenth-Century Journals Portal* (<http://www.amdigital.co.uk>) contains journals digitized from the Harry Ransom Humanities Research Center.
3. Members of these boards are listed at http://www.18thconnect.org/18th_about/boards.html
4. Jerome J. McGann, "Culture and Technology: The Way We Live Now, What Is to Be Done?," *New Literary History* 36, no. 1 (2005): 71–82.
5. This helpful analogy was drawn at a meeting of ARC, the Applied Research Consortium for digital futures, held at the University of Virginia by Andrew Stauffer, Director of NINES, 18–19 Sept. 2010.
6. The NEH-Funded Text Encoding Initiative developed a set of tags that are typically needed in scholarly and linguistic editing (<http://www.tei-c.org>), and the Guidelines for Vettors of Scholarly Editions at MLA asks reviewers, "Is the [document encoded] using relevant community guidelines (e.g., the Text Encoding Initiative guidelines)?" (http://www.mla.org/cse_guidelines#d0e354).
7. See JuXta (<http://www.juxtaoftware.org/>); the Versioning Machine (<http://www.v-machine.org/>); TAPor (<http://portal.tapor.ca/portal/portal/>); Voyeur: Reveal Your Texts (<http://www.hermeneuti.ca/voyeur/>); and TokenX (<http://jetson.unl.edu:8080/cocoon/tokenx/index.html?file=../xml/base.xml>).
8. In such cases, any images accompanying the scholarly edition could be collected in the traditional way by contacting the holding library, requesting permissions and photos for reproduction.
9. John Guillory, "Close Reading: Prologue and Epilogue," *ADE Bulletin* 149 (2010): 8–14.
10. *Graphs, Maps, and Trees: Abstract Models for a Literary History* (New York: Verso, 2005); Franco Moretti, "Style, Inc.: Reflections on Seven Thousand Titles (British Novels, 1740–1850)," *Critical Inquiry* 36 (2009): 134–58; Katie Trumpener, "Paratext and Genre System: A Response to Franco Moretti," *Critical Inquiry* 36 (2009): 159–71; Franco Moretti, "'Relatively Blunt,'" *Critical Inquiry* 36 (2009): 172–74.
11. John Guillory, "How Scholars Read," *ADE Bulletin* 146 (2008): 8–17.
12. Normally, for a full-text only search, you would see that constraint at the top of the advanced search page, but, when I conducted this experiment, we had only put full-text ECCO items from the Text Creation Partnership into 18thConnect:

Search Query (LOG IN to save this search) New Search
(LOG IN or Create new account to save these results)
Add new search criteria or select limiters to refine your search.

Full Text	Only resources that contain full text	AND	🗑️
Search Term	click here to add new search term	AND	Add

13. See the introductory film on 18thConnect.org: go to "What is 18thConnect?" (http://www.18thconnect.org/18th_about/what_is.html), available by clicking on that phrase on the home page (top right).
14. A recent essay in ProfHacker explains how. See Amy Earhart, "Using NINES Collex in the Classroom," *The Chronicle of Higher Education* ProfHacker Blog, <http://chronicle.com/blogs/profhacker/using-nines-collex-in-the-classroom/23829>.
15. See the Digging into Data Challenge home page at the NEH for more information: <http://www.diggingintodata.org>.