Critical Editing in the Digital Age: Informatics and Humanities Research

Eduardo Urbina, Richard Furuta, Arpita Goenka, Rajiv Kochumman, Eréndira Melgoza, Carlos Monroy

Center for the Study of Digital Libraries
Texas A&M University

Our project—The Electronic Variorum Edition of Don Quixote, 1605-2005 (EVE-DQ)—is a collaborative initiative started in 1998 by a team of researchers in Humanities and Computer Science at Texas A&M University’s Center for the Study of Digital Libraries. It received initial research support from the Office of the Vice President of Research and the College of Liberal Arts, and is funded at present by a three-year grant from the National Science Foundation programs in Digital Libraries and Information Technology Research. The resulting research and edition will be integrated into the Cervantes Digital Library, part of The Cervantes Project (CP), in time for and as a contribution to the celebration of Don Quixote’s fourth centenary in 2005.

The focus of the EVE-DQ is the application of advanced computer science research to update and enhance traditional scholarly practices. Our primary goal is to develop a replicable program that permits the creation of online critical editions as hypertextual archives, using the Quixote as testbed. In this context, and from the perspective of our specific experience, I will comment today on the state of critical editing, editors, and editions in the digital age and how it relates to the future of the tradition of bibliographical and textual studies as established by Greg, Bowers, and Tanselle. In particular, I will address and examine the impact of the Internet and information technology upon the three key scholarly practices in textual criticism: recensio, collatio and emendatio, and the new role of editors (and readers) in the creation of virtual electronic editions by direct access to online hypertextual archives and databases.

As far as the development, production, and publication of complex scholarly editions, such as variorum editions, the future of such texts as archives is most certainly electronic, networked, and hypertextual. Nowadays, it would be unthinkable and unfundable to propose to undertake in a traditional fashion a project such as The New Variorum Shakespeare—conceived in 1871 by Horace Howard Furness, and still to be completed—using the means and technologies associated with paper and the printing press. Perhaps equally clear, although not so obvious, is the notion that computer-assisted editions of a similar nature, that is to say, editions that aspire to produce printed editions using electronic tools, are also, to a large extent, wasteful and even...
anachronistic. Even hybrid editions such as the remarkable edition of the Quixote published in 1998 under the direction of Francisco Rico, which includes an electronic version of the text in an enclosed CD-ROM, plus a textual analysis tool and web support, do not fully take advantage of currently available technologies and advances in electronic editing, much less of the type of new developments in textual criticism recommended since the early 1990’s by Charles Faulhaber, George Landow, and Peter Robinson among others, conceptualized by Jerome McGann in his trailblazing study “The Rationale of Hypertext,” and exemplified in his project, The Rossetti Archive.

In the last few years we have seen the rapid appearance and proliferation of collections and web sites dedicated to the electronic conversion of previously published texts—by necessity not always the best, most critical or most recent--from those in the Gutenberg Project to some included in our own Cervantes Project since 1996. Almost none of these electronic text collections include search engines or tools for textual analysis, and some do not even provide documentation about the provenance of the texts being transcribed, the manner of conversion, not to mention textual notes or bibliographical references. The second generation of electronic texts exhibits a more critical approach to the conversion process and is primarily concerned with encoding and tag inclusion to be able to preserve and transmit electronically their original appearance and the formal characteristics of the printed page. A great deal of time and resources have been dedicated recently to insure that in their transition to electronic format printed texts retain the markings, format, and organization dictated merely by the conditions and technology under which they were first composed and published. Whether we decide to consider these efforts descriptive or prescriptive, I am of the opinion that we are dealing with a losing battle, and a very costly one besides. Although these interests and initiatives could and often are justified from an archival and scholarly perspective, they have sidetracked and delayed, I believe, the development of tools and programs to create pure electronic editions, that is to say, editions that from the start have as a goal the development of electronic texts using digital means and information technology research. These new forms of textuality aim to produce hypertexts, hypereditions, and hypertextual archives of the type anticipated by Landow, envisioned by Faulhaber, Marcos Marin, Hockey, and Lavagnino, and initiated by Robinson, Donaldson, and McGann. Our project, the EVE-DQ, aspires to be such a third generation electronic edition, as I shall describe shortly.

In addition to the evolutionary delays and constraints so far described, from first to third generation electronic texts and editions, a number of factors still impede and make difficult the wide development of hypereditions and hypertextual archives at the present time. Among them are: the lack of acceptable OCR programs, customizable and capable of rendering with a high degree of accuracy the early European typefaces characteristic of the hand-press period, as extant in the works of Shakespeare and
Cervantes. Most significantly, however, is the lack of a critical mass in the area of infrastructure research in computer science, digital libraries, and humanities informatics to facilitate the coordinated production of scholarly electronic hypereditions in the few existing Centers and Institutes dedicated to the application of information technology to traditional bibliographical and textual studies. This situation is exacerbated furthermore by a reluctance, still, to work outside traditional disciplines, thus perhaps putting at risk academic advancement and recognition, or the inherently difficult issue of being able to participate in collaborative interdisciplinary projects involving areas as distant from each other as textual studies and computer science. Finally, the organizations dedicated to the culture of the book—publishers, libraries, booksellers, etc.—have been hesitant for obvious economic and political reasons to surrender the instruments that empower them, or, at best, seem to feel a deep-seeded ambivalence regarding the certain threat and potential benefits represented by the new forms of textuality, particularly when associated with a free and universal distribution network such as the Internet. The future is indeed a most uncertain proposition and archives are, after all, stable and reassuring entities, human constructs and mechanisms, if you will, to control and perpetuate as past the transient nature of the present as realized future.

Given these universal factors and the levels of uncertainty and fear they generate when combined, it is not surprising that the number of theoretical and programmatic studies envisioning new types of texts and editions, or proposing new paradigms of electronic textuality far exceed the actual examples of developed projects and scholarly collections of critically edited literary electronic texts, freely accessible to all users. For now, third-generation digital collections of literary texts such as the works of Chaucer, Shakespeare and Cervantes remain complex, expensive, and difficult enterprises caught between tradition and innovation. They require for their successful production and existence a delicate combination of original collaborative scholarship, groundbreaking interdisciplinary research, and forward looking financial and institutional support, such as one can find and appreciate in The Canterbury Tales Project, the Shakespeare Electronic Archive and the Perseus Digital Library, to cite only some of the most notable collections and projects.12

In the context of assessing electronic editions with reference to Shakespeare and the Internet, R. G. Siemens was able to observe still in 1998 that not only “scholarly editing . . . appears to be a site of some contention” but goes on to state, most significantly for us, that “the salient features of the electronic edition remain unarticulated in the larger discourse of literary-textual culture.”13 Indeed, textual scholars and editors have been slow to react to the new electronic paradigm, and even more so in recognizing the sure impact and profound implications of the Internet on their discipline. Thus, even such pioneering minds and hypertext practitioners as Landow and Robinson were addressing in 1996 the technical problems posed by annotating hypertexts and the future direction of hyperediting.14 And it was not until
the 3rd edition of Shillingsburg’s *Scholarly Editing in the Computing Age* (1996), for instance, that a chapter on electronic editing was included, and even then it was infused by a serious dosis of skepticism as to the merits of non-linear editions.\(^\text{15}\)

In spite of all the hype and anticipation brought about by the digital revolution, along with the somewhat prematurely declared death of the editor and of traditional editing, the truth is that “electronic editions have been slow in appearing.” The existence of a mere dozen or so major electronic publishing projects, according to Robinson, seems to reveal precisely that “the power of the computer to search, to sort, to navigate, to explore” remains an unrealized promise, as his own proposal for a future model and methodology would confirm.\(^\text{16}\) Even most telling is the confessional statement made in 1992 by Richard Knowles regarding the then contemplated possibility of the MLA publishing the new *variiorum* Shakespeare both in book and electronic form. According to Knowles, “the program has not yet been invented that can provide access to information as well as the printed volume.”\(^\text{17}\) To this date, and to my knowledge, no such edition has been forthcoming. Nevertheless, Siemens, Cook, Hockey and others have dared to anticipate and predict the future of hypertextual archives and to envision, even, the coming of a new type of electronic editions capable of encompassing and providing access to all available textual knowledge and scholarship--graphical, critical, and historical. Our quixotic project, in all modesty, attempts to accomplish such a goal by developing an end-to-end program and tools capable of converting, collating, annotating, and composing, within the classical tradition of textual studies, multiple scholarly editions, from a documentary edition to a critical *variiorum* edition.

There has never been a textual or critical *variiorum* edition of *Don Quixote*, and in fact, according to Paul Wernstine, “there has never been an old-spelling edition of Shakespeare’s so called canon prepared according to the principles of copy-text editing . . . definitive and entirely objective.”\(^\text{18}\) Such impossible quest seems to have dissipated for now in the age of poststructuralist theory and textual deconstruction, as a fatal doubt was cast on the idea of reconstructing authorial meanings out of non-existent texts. In fact, the rise of multiple interpretative communities has contributed in part to the development of new editorial practices in the 1990’s which include, as a matter of fact, hypertext editions. Notably, as Wernstine observes again speaking of Shakespeare’s works,

Hypertext allows the Shakespearean editor finally to confront in practice the knowledge that since we do not know anything specific about the manuscript sources of the printed plays, we never had any grounds (besides aesthetic or practical ones) for choice between one printing of a play or another. . . . The challenge for editors shifts from the task of generating editorial theory that can
be deployed in the defense of archaic editorial choices to the practical task of how to display in book form for a wide audience the indeterminacy that characterizes Shakespeare textuality. (52)

Needless to say, the very same words and criteria can be applied to Cervantes’ works, and the *Quiçote* in particular, for which no manuscript sources of the printed text exist either, and with the same liberating effects. Our challenge, however, has been from the beginning how to make available for the first time the complete textual record to the widest audience in the form not of a cumbersome, expensive, multivolume printed book, but in the only manner feasible in the digital age, that is to say, in the form of a hypertextual *variorum* edition or archive.

The *EVE-DQ* is not, strictly speaking, an edition but rather a dynamic hypertextual archive composed of a series of databases—graphical, textual, and documentary—and two principal modules, the editing module or MVED (Multi-variant editor of documents), and the visualization-composition module or VERI (Virtual editing-reading interface), which in turn are controlled by a hypermedia-based data entity management system (HDEMS).19 The initial process of collation involves the creation of data entities. For example, a variant that is being annotated is a data entity. The HDEMS manages these data entities by preserving the relations and linkages among them, and maintaining consistencies amongst related data entities. The HDEMS API keeps track and manages within the MVED the relational databases created in the collation/editing/annotation/emendation process. All other modules call the HDEMS API functions and the actual interaction with the databases is done by the HDEMS API.

The editing program or MVED, is comprised of seven tools or modules, which can be briefly described as follows:

1. The *Collator program* automatically identifies textual variants among two or more of the selected editions/copies against a base text.
2. The *Index of Variants* generates multiple lists of textual variants identified during the collation and provides interactive visual representations of such variants.
3. The *Text-Image Synchronizer* uses predetermined synchronization points between the digital images and the electronic texts for each of the editions/copies included in the collation process.
4. The *Dual-form Document Viewer* shows together the synchronized texts and associated images to be able to browse the source editions incorporated into the collation.
5. The *Editing/Emendation tool* allows the editor to correct and annotate both variants and non-variant text, and creates the hyperlinked data entities stored in the relational database.
6. The *Data Entity Browsing Interface* provides the editor with updated access to all the
details and information produced during the editing process.

7. The *Annotation Tool* allows the classification and annotation of variants as well as the inclusion of critical commentary and references during the editing process of any of the collated texts.

The following figures show a number of windows and views illustrating the operation of the MVED modules during the collation process:

![Figure 1: MVED tools during a collation](image1)

![Figure 2: Text-image synchronizer, List of variants, and Annotation tool](image2)
Essentially the MVED is a collation tool. Modules such as the “Collator program,” the “Lists of variants,” the “Text-image synchronizer,” and the “Dual-form
document viewer” function within the collation process but are independent of the editor being logged in, while the “Data Entity Browsing Interface” is used only within the collation process. One can, however, browse through documents using the browsing interface. Thus, although all the tools are used in the collation process, the “Dual-form document viewer” and the “Text-image synchronizer” can be used independently. In addition, the “Editing/Emendation tool” and the “Annotation tool” are designed to be used together so that one can annotate while editing a variant, or while doing an emendation. This dual function is only possible during the collation process, and it is necessary that a valid editor be logged in to the system.

The EVE-DQ constitutes a new approach to traditional critical editing and textual studies in the digital age and a new direction in the application of advanced computer science research and information technology to traditional scholarly practices. As such, it represents a fundamental update to the three key operations of the editing process: recensio, collatio and emendatio.

Williams and Abbot define recensio as “the reconstruction of the lost common ancestor of surviving texts through the examination of manuscripts and the texts they contain....” It is clear that the digital environment of the EVE-DQ allows not just the examination and reconstruction of texts in a selective and exclusive fashion by a privileged and authoritative editor, but the inclusion and visualization in an electronic hypertextual archive of all the surviving copies and editions, both in image and textual format, for the examination of any editor or authoring user. This level of access to textual information in the form of digital facsimiles and documentary texts, until now impossible to imagine in textual studies, has as a consequence not only a more direct and universal way of gathering and sharing knowledge but implies a change in the rationale and function of the critical apparatus, the key element of any scholarly edition. On the one hand, such complete, universal, and immediate type of textual access makes superfluous to a large extent notes and other items describing the accidentals of the texts, while at the same time it obviates the need for many editorial choices, since the editor is no longer driven by the need to reconstruct and publish a unique or single edited text. Any reader can at anytime visualize the text of any particular copy or edition without the need to accept the hidden choices or personal conjectures of the editor. The new paradigm of textual editing thus created by the new technology profoundly affects non only the relationship between the text (author) and editor but the role of the editor as agent of knowledge, and the role of the reader, no longer a passive receptor of filtered or absent documentation, but instead a fully informed and active participant.

In textual criticism, collatio is the operation by which one text is compared to another text to discover textual variation. This time consuming and tedious process has been the object in the past of several automation attempts, from Hinman to
Schillinsburg to Robinson, and has then benefited most from technology and computer applications. The MVED as a collation tool, along with the seven editing modules it includes, represents clear evidence of the new paradigm of textual studies. As described above, the MVED not only allows the collation of two or more texts but offers a complete flexible and interactive index of variants, the ability to automatically create links from the variants to the texts, and to synchronize them with the digitized images of the textual sources. In addition, it offers a mechanism to edit and annotate variants and non-variant text, and to store all annotations, commentaries and references in a relational database for future use by another editor, or to be included in a customized virtual edition by an individual user through the VERI. Again, such level of access and flexibility makes unnecessary, and perhaps obsolete, many of silent editorial choices and eliminations imposed by the technological and economic limitations of the printed edition in book format, starting of course with the controlling fixed concept of the “copy-text” and the illusory notion of the “definitive” edition.

In classical textual criticism, *emendatio* is “the use of conjecture to emend the text reconstructed from surviving witnesses, thereby removing errors.” In the digital context and through the editorial practices that characterize the MVED and our EVE-DQ, the need for *emendatio* if not altogether eliminated, becomes greatly diminished. Since the number and types of conjectures introduced by the editor are often driven by a limited collation process and predicated by the traditional methods of the *recensio*, it is clear that the paradigmatic effects caused in such critical processes by the new technology and tools also have profound consequences in the new electronic edition. The type and quantity of textual emendations needed in the open, flexible, and interactive environment of the hypertextual archive is greatly reduced, being replaced instead by a more transparent and inclusive access to the complete textual knowledge it affords.

To conclude, the future of the archive--of the textual electronic archive—will increasingly take the form of hypertextual editions or hypereditions, such as the EVE-DQ, and constitute already a reality impossible to ignore. This new paradigm in textual studies, and the system and tools it provides, is rapidly changing the way we read and the way we learn, the way we work and the way we teach the texts and documents that inform the Humanities as a discipline, in general, and literary studies in particular. The digital revolution, as did the Gutenberg revolution, has already changed the way we communicate and the way we live. It will be naïve to think that such pervasive and revolutionary changes will not equally affect the nature and future of the archive and of its most traditional embodiment, the scholarly critical edition, as the premier humanistic instrument of philological, textual, and digital knowledge.*

---

* This material is based upon work supported by the National Science Foundation under Grant No. IIS-0081420. Support for this work also was provided by the Interdisciplinary Research Initiative Program, administered by the Office of the Vice President for Research, Texas A&M University.
Notes


5 El ingenioso hidalgo don Quijote de la Mancha, Dir. Francisco Rico. 2 vols. (Barcelona: Instituto Cervantes-Critica, 1998. Over 100 contributors participated in Rico’s edition, and it was sponsored by the Centro para la Edición de los Clásicos Españoles. It has been revised and reprinted several times already and the text and notes are now available online: http://cvc.cervantes.es/obref/quieto/.


8 For such a collection and texts in the context of Spanish literature, see the *Biblioteca Virtual Miguel de Cervantes*, [http://cervantesvirtual.com/index.shtml](http://cervantesvirtual.com/index.shtml).


17 Quoted by Cook, see note 4 above.


20 The concepts and definitions here used are taken from William Proctor Williams & Craig S. Abbott, *An Introduction to Bibliographical and Textual Studies*, 3ª ed. (New York: The MLA, 1999).