Towards an Electronic Variorum Edition of *Don Quixote*

Richard Furuta, Shueh-Cheng Hu, Siddarth Kalasapur, Rajiv Kochumman, Eduardo Urbina,
Ricardo Vivancos-Pérez[1]

Center for the Study of Digital Libraries
Texas A&M University
College Station, TX 77843-3112, USA
{furuta, shuehu, ssk9770, rajiv, e-urbina, rv} @csdl.tamu.edu

ABSTRACT
The Cervantes Project is creating an Electronic Variorum Edition of Cervantes' well-known *Don Quixote*. This paper gives an overview of the computer-based tools that we are using in this endeavor, and summarizes the current status of the project. The Electronic Variorum Edition will join the other content elements maintained by the project, which focuses on electronic resources in support of the study of Cervantes, his works, and his times.

Keywords
Humanities digital libraries, Hispanic culture, Cervantes Project, Cervantes Digital Library (CDL)

1. INTRODUCTION
The Cervantes Project, housed under the auspices of the Center for the Study of Digital Libraries at Texas A&M University, seeks to provide a comprehensive on-line reference and research site on the life and contributions of Miguel de Cervantes (1547-1616), the author of the classic *Don Quixote de la Mancha*. The project, initiated in 1995, contains a number of components: the Cervantes Digital Library (CDL), with copies of electronic editions of Cervantes' novels, plays, and other related writings; the Cervantes Digital Archive of Images (CDAI), a developing archive of photographic images on Cervantes' times and places suitable for teaching and research purposes; and the Cervantes International Bibliography Online (CIBO), a comprehensive bibliography of studies, editions, and translations of Cervantes' works.

Currently, we are creating an Electronic Variorum Edition of *Don Quixote* (EVE) for inclusion in the CDL. The EVE will contain all of the significant early editions of the text in interlinked textual and image form. In addition, it will support scholarly analysis of the differences between the editions along with interpretative commentary. The reader of the EVE will be able to customize their view of the text, examining and selecting among (or possibly combining) the choices made by independent editors. Original source material will be available to the reader, as well.

Don Quixote was published in Madrid in two volumes—1605 and 1615 (these editions are called the *princeps*). The original manuscripts have not survived and indeed only 18 copies are known to exist of the 1605 *princeps*. Consequently, our work centers around the significant editions printed during Cervantes' life. Including the two *princeps*, there are five such editions for volume 1, two for volume 2, and two containing both in a single book. Each volume is about 700 pages in length. To date we have obtained and scanned five microfilmed copies each of the two *princeps* and one or two copies of the remaining editions, for a total currently on-hand of 21 copies.

Our work-plan for each volume is first to create an "ideal" version (a *base text*) of the *princeps* by consulting with multiple printings, identifying differences in the individual copies, and then making reasoned judgments as to a preferred rendition; this process is already underway for volume 1. To the extent permissible by our use agreements with the libraries that own the copies, we also plan to find the best available image for each page of those available to us. Subsequently, we will collate the base text against the other editions to produce the *documentary variorum edition*.

Editors of an EVE need to collate editions to find differences, resolve the differences among variants, and provide additional annotation; this is supported by the MVED (the multi-variant document editor). Readers of an EVE need the abilities to examine and customize editions; this is supported by a separate Reader's Interface. Earlier versions of these interfaces have been described elsewhere [2]; here we will sketch some of the issues we have encountered.

2. COMPONENTS OF THE EVE
2.1 Microfilms and Images
We are producing the EVE from scanned microfilm images, both for the practical reasons of access and expense, but also to investigate what could be achieved with available-quality images (our collection currently greatly exceeds what is generally available to the Cervantes scholar). Figure 1 shows a representative image.

The straightforward processing will be to "trim" the images to remove background and to correct imaging skew. This process already is underway. We also are evaluating transformations that might be applied to the text area. Here, we must take a cautious approach; see for example Donaldson's description of semantic differences introduced by interpretation of an ambiguously-printed character as "f" or "s" [1]. The clear implication here is

[1]The authors are listed alphabetically. Richard Furuta, Siddarth S. Kalasapur and Rajiv Kochumman also are affiliated with the Department of Computer Science. Eduardo Urbina and Ricardo Vivancos-Pérez also are affiliated with the Department of Modern and Classical Languages. The project's Web pages are at http://www.csdl.tamu.edu/cervantes/
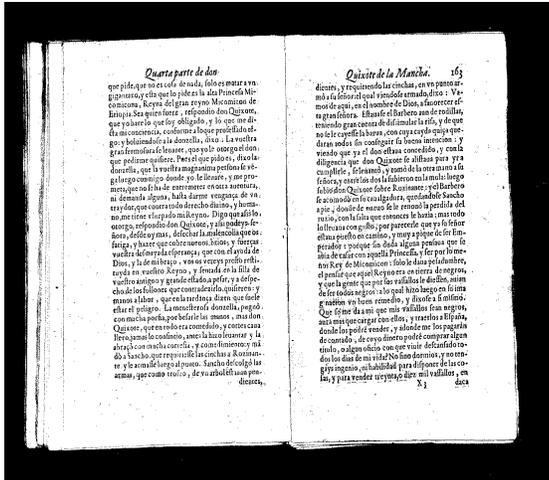
Figure 1: A representative page from microfilm.

that for major "corrections", both cleaned and original images must remain available for inspection by editors and readers.

2.2 MVED

The MVED is our software tool that enables scholars to identify, analyze, and edit variances in collated texts, given a chosen base text and different editions of the same. The MVED also enables scholars to select and annotate sections of the text.

The MVED's collator module automatically identifies variances between the base and one or more other texts. The scholar can classify acceptable variances, and can also identify variances not brought out by the collator module. Figure 2 shows a collation in progress using the MVED.

In order to aid the scholar in working with multiple texts, the MVED has a dual-form document viewer, which allows the editor to view the synchronized image of the actual document, along with the corresponding textual transcription.

Evaluation of the MVED has helped us better understand the role of annotations and the distinctions between the uses of annotation-like features (see also [3]). While variances require justification, a separate mechanism is needed to allow commentary. Additionally, the classifications appropriate for justifications are different from those appropriate for classifications. We currently support variants and annotations as separate mechanisms, each associated with free-text commentary. Additionally, as an accelerator, annotations can be attached directly to variants in addition to selected portions of the base text.

2.3 Readers' Interface

The Readers' Interface enables users to view the result of collations via the World-Wide Web. The MVED potentially makes a very large amount of data available to users, from the base texts and the comparison texts, to the collation results and statistics, along with various editors' comments on the same.

Our initial design adopted what might be called a "category-centric" model of desirable customizations (e.g., selection by category of variance, for example). Use of the interface by Humanities scholars pointed out the need instead to adopt an editor-centric model for customizations, as comprehensibility is closely tied to issues of authority.
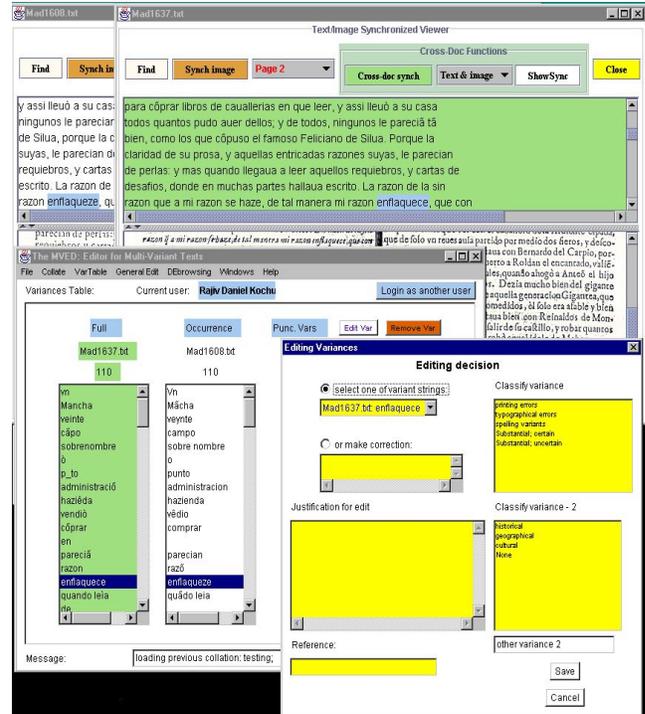

Figure 2: A collation in progress using the MVED.

3. FUTURE WORK

As our work continues, and we continue to gain experience with our potential users, additional requirements are identified that lead to rethinkings of our underlying models. A strong characteristic of our use environment is the need to support both English-language and also Spanish-language speakers—both editors and also readers. Consequently, the language, or languages, understood and preferred lead to further generalizations of the interrelationships among components in our system's architecture. Additionally, the broad popularity of Cervantes, and of *Don Quixote*, raise questions of appropriately supporting quite different categories of readers—ranging from University-level researchers to grade-school children.

4. ACKNOWLEDGMENTS

5. REFERENCES

[1] Peter S. Donaldson, *Digital Archive as Expanded Text: Shakespeare and Electronic Textuality*. In Katheryn Sutherland, editor, *Electronic Text: Investigations in Method and Theory*. Oxford Clarendon Press, 1997, pp. 173–197.

[2] Shueh-Cheng Hu, Richard Furuta, and Eduardo Urbina, "An Electronic Edition of Don Quixote for Humanities Scholars". *Document Numérique,* 3(1–2), June 1999, pp. 75–91.

[3] William Proctor Williams and Craig S. Abbott, *An Introduction to Bibliographical and Textual Studies*, Third Edition. Modern Language Association of America, 1999, pages 69–125.