

# ***The use of multimedia to enhance the accessibility of digital library resources***

## ***The multicultural-scope of the services offered by the Miguel de Cervantes digital library***

**Alejandro Bia**

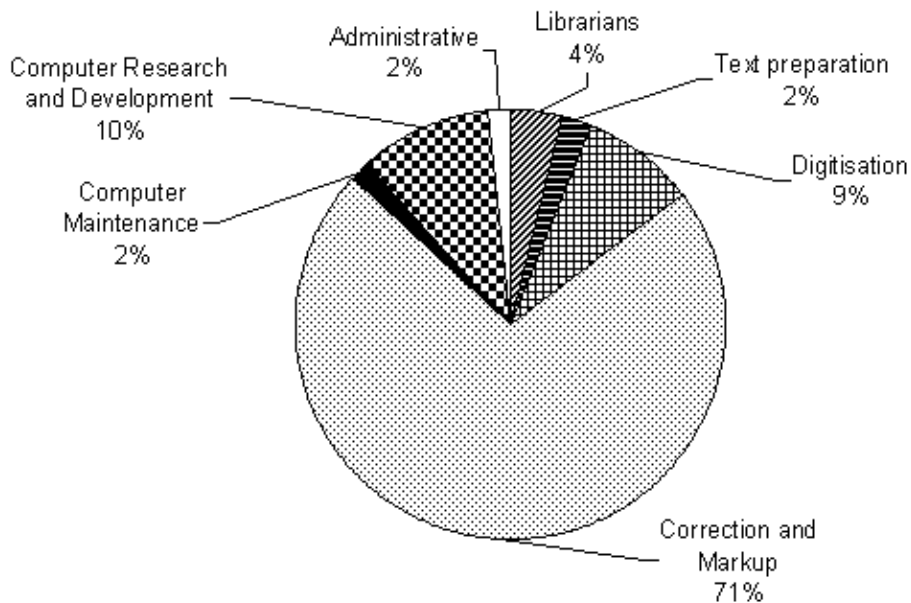
### **Description of the project**

The Miguel de Cervantes Digital Library of the University of Alicante, Spain (<http://cervantesvirtual.com>), covers many different cultural and academic subjects. It consists of a variety of types of multimedia materials such as hypertext books, facsimile books, images, audio recordings, and videos. Since the beginning of this digital library two years ago, about 64,000,000 accesses have been made. Currently, the average number of pages served is 21,000 daily. 71% of our users are from Spanish speaking countries (Spain and Latin America), 6% from the rest of Europe, and 23% from the rest of the world.

### **History**

The project began in late 1998 with a small team of 5 persons performing preliminary trial and error tests to determine the viability of the project and to find the most adequate production methods and tools. After this initial phase, massive digital book production started in March 1999, with about 80 persons working on digitisation and correction of texts and a small team of computer technicians mounting the computer infrastructure, developing automatic tools for efficient production and designing the publication processes and the web site. The library was opened to the Internet public in July 1999 with approximately 800 books. Now it has more than 6000 works in catalogue. An average of 120 university graduates and technicians is currently working in this project as librarians, scanner operators, correctors, mark-up specialists and computer technicians.

## Human Resources: 129 persons



### Financial support

The Miguel de Cervantes Digital Library is the result of a unique collaboration between the University of Alicante and Spain's biggest bank, the Banco Santander Central Hispano, who have joined forces to create the world's biggest digital library containing Spanish works. It represents an example of a successful partnership between university and business, with the Santander Central Hispano Bank providing complete sponsorship for the full development of the project. The University, on the other hand, provides the academic expertise, technological know-how and qualified workforce necessary to fulfil the objectives and ensures the international use of the Library's resources by means of collaboration agreements with universities and institutions all over the world [Bia and Pedreño, 2001].

### Multilingualism and multiculturalism

The library collection focuses on Hispanic literature, history and culture, in various Hispanic languages such as Catalan, Basque and Galician, or Latin-American native languages as Quechua, Mapuche or Guaraní in addition to Spanish.

It is meant for a wide audience covering different ages, interests and cultures. The library includes such diverse resources as digitised cultural magazines<sup>1</sup>, a library for children<sup>2</sup>, and academic research resources like PhD theses<sup>3</sup>.

<sup>1</sup> <http://www.cervantesvirtual.com/hemeroteca/>

<sup>2</sup> <http://cervantesvirtual.com/portal/platero/>

<sup>3</sup> [http://www.cervantesvirtual.com/tesis/tesis\\_catalogo.shtml](http://www.cervantesvirtual.com/tesis/tesis_catalogo.shtml)

Besides, the digital library has whole sections devoted to many renowned academic institutions and national libraries of Latin America, who contribute their valuable funds to our library for digitisation<sup>4</sup>. These sections are meant to preserve and promote minority Hispanic-related languages, and to affirm different Hispanic cultures. The design of these sections does not only entail the use of new techniques, but also involves collaboration agreements whereby staff from the Miguel de Cervantes Digital Library travel to other centres, universities and institutions all around different areas of Spain and Latin America in order to digitise the documents, objects and works of art that will be finally published on the library main server. After some training, staff from the contributing institutions begin to digitise and correct the texts which are finally marked-up, supervised, processed and published in Spain.

Existing multicultural sections include: the Library of Catalunya<sup>5</sup> and the Joan Luis Vives Institute<sup>6</sup>, both with works in Catalan language, country specific collections as the National Library of Chile<sup>7</sup>, the National Library of Argentina<sup>8</sup>, the College of Mexico<sup>9</sup> and the José A. Portuondo Valdor Institute of Literature and Linguistics of Cuba<sup>10</sup>, and thematic portals such as DOXA<sup>11</sup>, about the philosophy of law, Hispano-American Novel of the XIXth Century<sup>12</sup> (Benito Varela Jàcome fund), XIXth Century Spanish Literary Society<sup>13</sup>, Association of Spanish Theatre Authors<sup>14</sup>, Hispano-American Constitutions<sup>15</sup>, Geographic Research<sup>16</sup>, and a portal on Industrial and Intellectual Property<sup>17</sup>.

### **Libraries within the library dedicated to selected authors**

Other important sections of this digital library, under the title of “Authors' Libraries”, are special collections dedicated to selected world-renowned Hispanic authors<sup>18</sup>. They consist of personalized web pages describing their life and works with links to their digital books within the library as well as to external resources. Manuscripts and special valuable documents of the authors (like drawings, photographs, book covers and personal letters) are usually provided. In the case of contemporary authors, videos and audio recordings give a closer appreciation of their personality<sup>19</sup>.

---

<sup>4</sup> <http://www.cervantesvirtual.com/portales/portales.shtml>

<sup>5</sup> <http://cervantesvirtual.com/portal/BC/>

<sup>6</sup> <http://lluisvives.com/>

<sup>7</sup> <http://cervantesvirtual.com/portal/BNC/>

<sup>8</sup> <http://cervantesvirtual.com/portal/BNA/>

<sup>9</sup> <http://cervantesvirtual.com/portal/ECM/index.html>

<sup>10</sup> <http://cervantesvirtual.com/portal/ILL/>

<sup>11</sup> <http://cervantesvirtual.com/portal/DOXA/>

<sup>12</sup> <http://cervantesvirtual.com/portal/BVJ/>

<sup>13</sup> <http://cervantesvirtual.com/portal/SLESXIX/>

<sup>14</sup> <http://cervantesvirtual.com/portal/aat/>

<sup>15</sup> <http://cervantesvirtual.com/portal/constituciones/>

<sup>16</sup> <http://cervantesvirtual.com/portal/IIGG/>

<sup>17</sup> <http://www.uaipit.com>

<sup>18</sup> [http://www.cervantesvirtual.com/biblioteca\\_autor.shtml](http://www.cervantesvirtual.com/biblioteca_autor.shtml)

<sup>19</sup> [http://www.cervantesvirtual.com/bib\\_autor/mbenedetti/video.shtml](http://www.cervantesvirtual.com/bib_autor/mbenedetti/video.shtml)

Apart from this vast number of local resources, the digital library functions as a gateway to external Hispanic resources through a section called Libraries of the World<sup>20</sup> with links to other outstanding digital resources on the Web.

## **A wide range of media and a wide range of accessibility options**

### **A classification of digital objects according to multimediality**

According to the ease of transmission and computer requirements we classify our digital resources into six classes, from the more easily accessible to the more demanding ones.

1. hypertext books,
2. digital still images,
3. digital facsimiles with or without transcriptions,
4. digital audio recordings,
5. video plus text ensembles (signs library),
6. digital video recordings.

### **Hypertext books**

About 5,000 hypertext books were created and made available in the digital library. They are basically public domain Hispanic classics written or published between the twelfth century and the nineteenth century. Contemporary works are also included. Genres of these books include literary prose, drama, verse, dictionaries, historical documents, and others. The digital forms of these books were encoded by using the TEI (Text Encoding Initiative) standard [Sperberg-McQueen and Burnard, 1994] [Burnard, 1995]. This digital texts, which are the vast majority of the library works, are easily accessible for all kinds of internet user, even those with modest PCs and dialled-up connection.

Text offers both the advantage of being light in terms of binary size and transmission speed, while offering the versatility of electronic search, easy transformation to different formats with different rendering qualities, and the possibility of offering useful computer based services like automatic concordances.

The library provides both a catalog search, i.e. a search on metadata, and a book content search facility. For the latter, the indexation of the texts and the use of a conventional web search engine installed on our server allows for this service. Search over content can only be provided for text, while catalog search is provided for all six types of digital objects.

---

<sup>20</sup> [http://cervantesvirtual.com/bib\\_mundo/biblioteca\\_mundo.shtml](http://cervantesvirtual.com/bib_mundo/biblioteca_mundo.shtml)

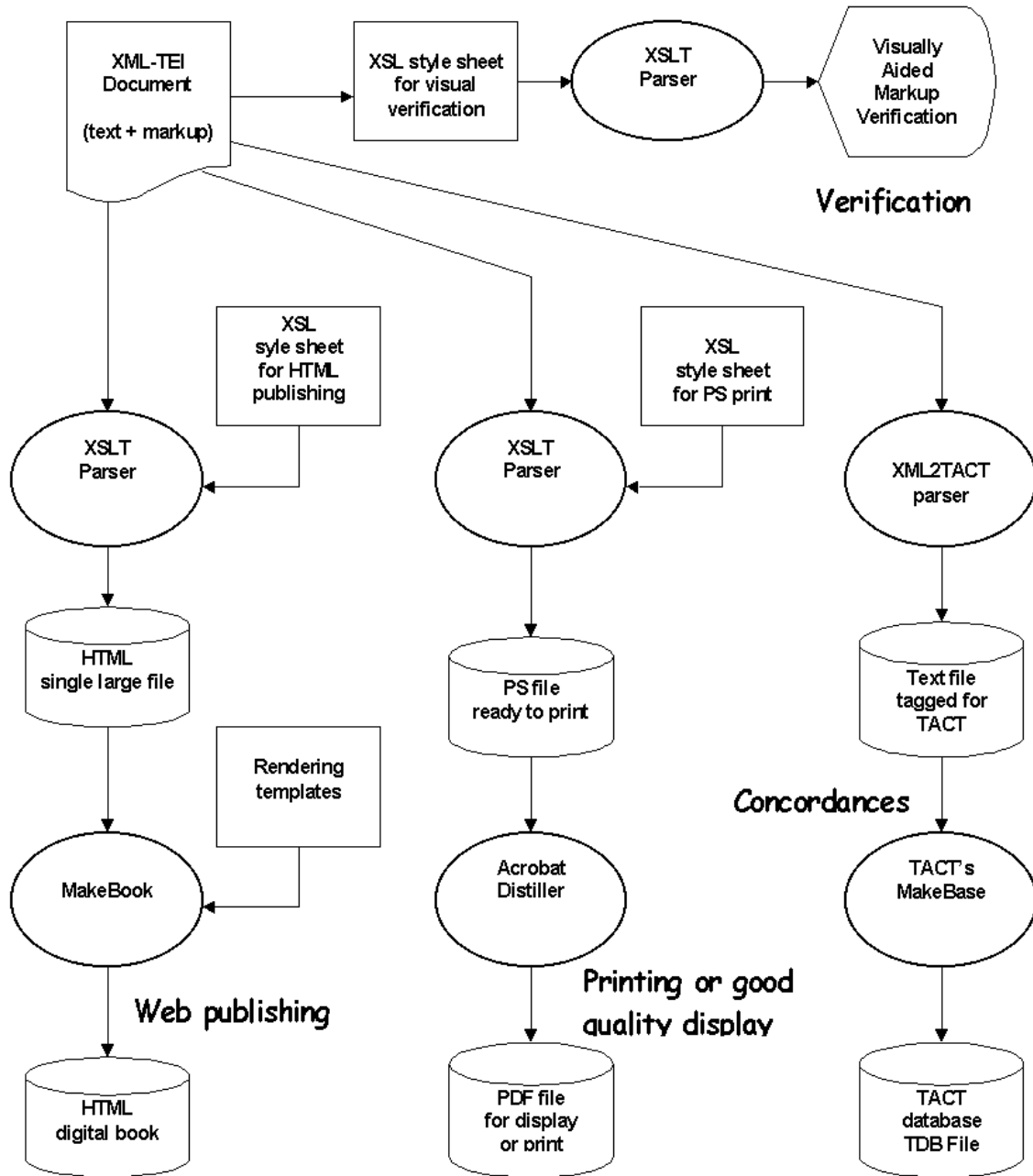
For the search in content we use HT-Dig, a conventional web-search engine available for Linux machines. A spider program periodically scans the contents of our library to update the search-in-contents index. For catalog searches we developed our own search engine, based on an object-oriented database and a fast access TRIE structure built in memory containing the necessary metadata [Bia and Nieto, 2000].

Concordances of selected works based on TACT [Bradley, 1991] and TACTweb [Bradley et al., 1997] are currently provided to be used by literary students and researchers<sup>21</sup>. TACTweb is a client-server-type interactive web service based on HTML form queries on the client side and a CGI program (TACTweb) that process the queries on the server side. TACTweb work on TACT generated databases. TACT is famous MS-DOS program for text concordancing and analysis. Both these programs were developed at the University of Toronto, Canada. Tags as used by TACT are represented in a manner similar to that used in COCOA markup. We developed a translation parser (XML2TACT) that automatically converts our TEI-XML texts to the COCOA-like input format used by TACT [Bia, 2000]. In this way, the generation of TACT databases from our XML texts is fully automatic and costless.

We also developed our own concordancing program for CD-based concordances [San-Martín et al., 2001]. In this case a client-server approach was not possible so we needed to develop a concordance service based only on HTML static files prerecorded on a CD-ROM. Our solution offers an HTML interface based on frames, where the user can select individual words and see their use in the selected text. It lacks the power of regular expression metacharacters (e.g. \* and ?) for which an intelligent software engine is necessary, but it provides a very useful web-independent service to students and researchers.

---

<sup>21</sup> <http://cervantesvirtual.com/FichaObra.html?portal=0&Ref=003603>



Digital hypertext-book production using XML-TEI and XSLT.

## **Enhancing accessibility: digital facsimiles**

At the same time, in the case of manuscripts and rare old books, facsimile versions of digital books were also produced<sup>22</sup>. Facsimile images have higher transmission demands, but are incomparable when dealing with manuscripts or outstanding old printings. They give distant readers the opportunity to see these valuable pieces without having to travel and ask for special access authorizations.

For this purposes, we developed an easy-to-use interface design that can provide facsimile images of pages in several sizes and formats (thumbnail, small page, big page or partial frames) together with up to three kinds of transcription (Madison style, normal, and modernized spelling). The user is allowed to combine this elements at will.

We took several measures to automate the production of this facsimile ensembles and reduce time and costs. On one hand, book pages are scanned once, and then automatically resized and cropped using a batch image processor, to obtain the different image formats needed for the ensembles. On the other hand, we developed a program (Facs-Builder) for massive and uniform production of these books that joins the images and transcription texts in an HTML set with the navigation functionality above mentioned [García:2001].

The book "El divino cazador"<sup>23</sup> is an example of the most complex design<sup>24</sup>, combining images of many sizes and different types of transcription (see figure). Simpler designs can be seen at our portal for the Biblioteca Nacional de Chile and the pages about Charles V<sup>th</sup> of our Library of History<sup>25</sup>.

---

<sup>22</sup> [http://www.cervantesvirtual.com/bib\\_autor/Calderon/manuscrito/p1.html](http://www.cervantesvirtual.com/bib_autor/Calderon/manuscrito/p1.html)

<sup>23</sup> [http://cervantesvirtual.com/bib/\\_autor/Calderon/manuscrito/p1.html](http://cervantesvirtual.com/bib/_autor/Calderon/manuscrito/p1.html)

<sup>24</sup> design by Francisco Perez

<sup>25</sup> <http://cervantesvirtual.com/historia/CarlosV/>

El Divino Cazador - Microsoft Internet Explorer

Archivo Edición Ver Favoritos Herramientas Ayuda Vínculos W M TABB Cines CV UA Google Intranet SPROD

Atrás Adelante Detener Actualizar Inicio Búsqueda Favoritos Historial Correo Imprimir Modificar Discutir

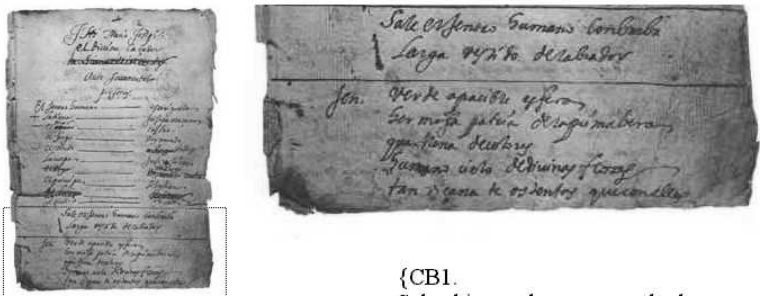
Dirección http://sprod/bib\_autor/Calderon/manuscrito/p1.html

BIBLIOTECA VIRTUAL MIGUEL DE CERVANTES

# El Divino Cazador

Pedro Calderón de la Barca

[Imagen completa](#) [Texto completo](#) [Transcripción Madison](#) [Transcripción](#) [Actualización ortográfica](#)



fol. 1v

{CBI.  
Sale el jenero humano con barba  
larga vestido de labrador  
jen<ero humano>. verde apacible  
esfera  
hermosa patria dela primavera  
que llena de colores  
humano cielo de diuinas flores  
tan viçarra te ostentas quecon ellas}

## Enhancing accessibility: digital audio

The audio collection features a Library of Voices with quality recordings of top Hispanic voices by volunteers<sup>26</sup>. Well-known Hispanic poets and authors also take part in the production by reciting parts of their own works. This collection is of interest for the general public, but specially for visually impaired users. A special interface using big letter size has been designed for this purpose.

## Enhancing accessibility: a talking interface

Additionally, an alternative navigation structure based on sound has been designed for the whole library in cooperation with ONCE (Organización Nacional de Ciegos de España: Spanish national organization for the blind). Blind users use a screen reader, a program that reads the contents of the screen using a synthesized voice, to navigate the library. There are various of these screen readers available nowadays, but our preferred one is a comercial program named JAWS. The user can use the reader not only to navigate the library, but to make it read entire book chapters. The voice of these readers although understandable, lacks the rythm, emotion and warmness of human recordings. Our Library of Voices is a preferable alternative in this sense.

<sup>26</sup> [http://www.cervantesvirtual.com/bib\\_voces/bibvoces.shtml](http://www.cervantesvirtual.com/bib_voces/bibvoces.shtml)



## **Enhancing accessibility: Signs language**

Our collection also features a Library of Signs<sup>27</sup> for serving hearing impaired users. Research and development effort is being applied on this project, where digital videos of persons “reading” a literary text using sign language are synchronized with the scrolling literary text itself. The texts used are specifically adapted from their originals for this purpose. The aim is to help the hearing impaired to associate the signs with the written language, which they find difficult to handle since they lack the ability to associate text with sounds.

A dictionary of signs is also provided for some selected words of the texts. By clicking on a highlighted word a pop-up video appears with the sign-language explanation of the selected word.

## **Interesting but ... less easily accessible: the video collection**

At the highest-bandwidth end of our multimedia classification, the video collection features recordings of cultural events, interviews, readings by original authors, comedy groups and selected parts of theatre performances.

We all know the power of audiovisual material for cultural or academic purposes but unfortunately video is a high bandwidth consumer and requires a fast computer with a fast connection to work properly. These requirements are not fulfilled by all of our users. Those with older machines and slow modem connections from distant locations are not yet able to see our videos. However, the rate of growth of the Internet and the trends of computer technology makes the use of video a sure bet for a near future, being the standardization of video formats, of which there are too many, a problem left to be solved.

## **Communication and interactive services**

With the aim of giving something close to the personalised service of a conventional library, the digital library provides communication facilities such as a virtual librarian, debate forums<sup>28</sup>, our own electronic news bulletin<sup>29</sup>, and a service devoted to the exchange of information to locate rare and out-of-print editions of books between readers, collectors, publishers, and booksellers<sup>30</sup>. The Virtual Librarian is a web-form based service, made available to answer standard questions<sup>31</sup>.

Most of the library deals with either public domain ancient books, or contemporary works whose copyright have been donated to the library. To cover the space of contemporary

---

<sup>27</sup> <http://cervantesvirtual.com/portal/signos/>

<sup>28</sup> <http://www.cervantesvirtual.com/foros/foros.shtml>

<sup>29</sup> [http://www.cervantesvirtual.com/bol\\_noticias/noticias.shtml#Dulcinea](http://www.cervantesvirtual.com/bol_noticias/noticias.shtml#Dulcinea)

<sup>30</sup> <http://www.cervantesvirtual.com/trueque/Trueque.shtml>

<sup>31</sup> <http://www.cervantesvirtual.com/bibliotecario/>

commercial copyrighted works, the library has developed a section called Primera Vista<sup>32</sup> ("First Sight"), which is a service of book reviews and previews of the latest paper editions, where first chapters of the books can be read on the Internet.

In this way, the **Miguel de Cervantes Digital Library** aims to become the most ambitious project of digitisation of Spanish works through the application of the newest methods of electronic publication to reach a massive multicultural audience of Spanish speakers.

## References

[Bia, 2000] Bia, A. (2000) XML2TACT: a translation parser to convert TEI-XML texts to TACT COCOA-like input files.

[Bia and Nieto, 2000] Bia, A., Nieto, A. (2000) Information Retrieval in Digital Libraries: efficient catalog searches using tries. Technical Report 2000-01, Miguel de Cervantes Digital Library, University of Alicante, Alicante, Spain.

[Bia and Pedreño, 2001] Bia, A. and Pedreño, A. (2001). The Miguel de Cervantes Digital Library: The Hispanic Voice on the WEB. LLC (Literary and Linguistic Computing) journal, Oxford University Press, 16(1):149-166. Presented at ALLC/ACH 2000, The Joint International Conference of the Association for Literary and Linguistic Computing and the Association for Computers and the Humanities, 21/25 July 2000, University of Glasgow.

[Bradley, 1991] Bradley, J. (1991). TACT Design. In Wooldridge, T. R., editor, A TACT Exemplar, volume 1 of CCH Working Papers, pages 7-14. Centre for Computing in the Humanities, Toronto.

[Bradley et al., 1997] Bradley, J., Rockwell, G., and Stevens, A. (1997). TACTweb, an experimental software to access TACT databases through Internet. <http://tactweb.humanities.mcmaster.ca/index.htm>.

[Burnard, 1995] Burnard, L. (1995). Text encoding for information interchange: An introduction to the text encoding initiative. <http://www-tei.uic.edu/orgs/tei/info/teij31/index.html>.

[García:2001] García, J.C. (2001) Facs-Builder: a program to build hypermedia facsimile ensembles. Technical Report 2001-02, Miguel de Cervantes Digital Library, University of Alicante, Alicante, Spain (2001)

[San-Martín et al., 2001] San-Martín, I., Pomares, J., Bia A. (2001) CD-Concord: a concordance generator based on HTML and JAVA. Technical Report 2001-03, Miguel de Cervantes Digital Library, University of Alicante, Alicante, Spain (2001)

---

<sup>32</sup> <http://cervantesvirtual.com/pvista/PrimeraVista.shtml>

[Sperberg-McQueen and Burnard, 1994] Sperberg-McQueen, C.M. and Burnard, L., editors (1994). Guidelines for Electronic Text Encoding and Interchange (Text Encoding Initiative P3), Revised Reprint, Oxford, May 1999. TEI P3 Text Encoding Initiative, Chicago - Oxford.