

Automating the internal workflow of the



<http://cervantesvirtual.com/>

poster by:

Alejandro Bia

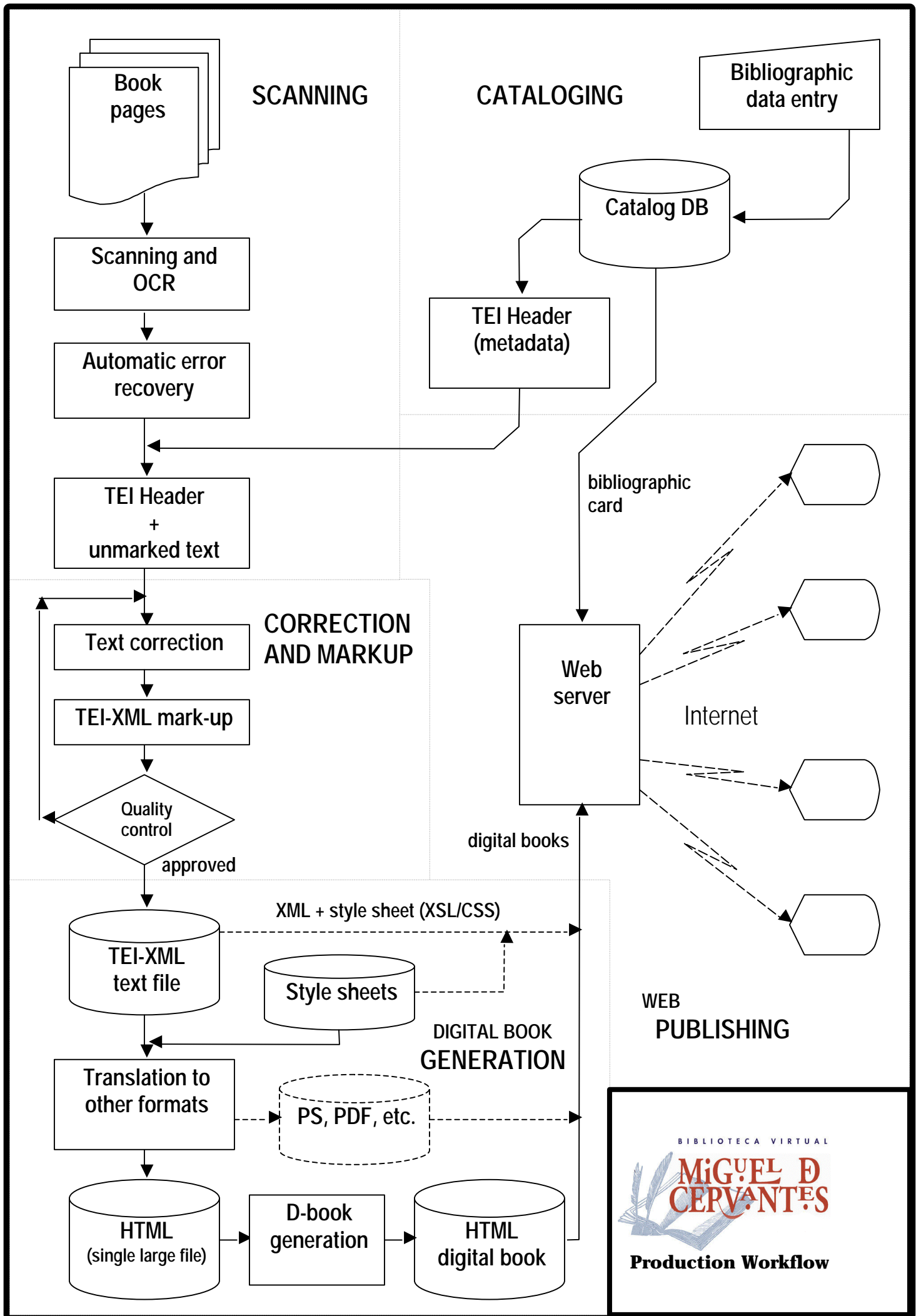
E-mail: abia@dlsi.ua.es

Biblioteca Virtual Miguel de Cervantes Saavedra

Universidad de Alicante, Alicante E-03080, Spain

<http://www.dlsi.ua.es/~abia/>

Tel: 3 4-965909567



Introduction:

Our library covers many different areas, from a “library of voices” up to academic thesis.

However, the vast majority of our present digital books are public domain hispanic classics, from the 12th century up to these days, including narrative, theater, poetry, history and other subjects. Many professionals and technicians take part in the development of our digital books: librarians, scanner operators, correctors, markup specialists and computer technicians. The diagram describes the production process of digital books and the information system that supports it, which we call “the workflow system”.

Cataloging:

The production process begins with a bibliographic search to find interesting available books to digitize. After selecting new literary works to add to the collection, the librarians elaborate the orders to be sent to various sources (conventional libraries, bookstores, publishers, private collectors in the case of rare books, etc.).

Upon reception, the books are cataloged. Information like subject, authors and collaborators, universal decimal classification and search keys that will simplify the location and retrieval of the books is stored in a database.

Scanning stage

The resulting output of the scanning process is a set of files of two classes: first, scanned images, and then optical character recognition (OCR), text documents. The former are stored in backup media for future projects. The latter, after an automatic error recovery process, are passed to the correction stage.

Correction:

The next stage is correction, where specialists in literature, history and linguistics use a text editor to correct not only OCR errors but also mistakes in the original publication, sometimes making side-by-side comparisons among different editions.

Markup:

Markup, consists of applying marks that describe the structure of the documents, and also marks that indicate how the digital book should be built later. These marks are concerned with index inclusion, chapter fragmentation, hypertext linking and graphics insertion, among others.

After markup, documents undergo a final revision before being passed to the following stages.

Generation and publishing:

Finally, the digital book generation stage is a fully automated process where marked-up document files are processed by a parsing program that produces HTML digital book files ready for immediate Internet publishing.