The use of computers for the purpose of facilitating the work of compiling the Wisconsin Dictionary of the Old Spanish Language (DOSL) is still, relatively speaking, in its cradle years. Yet, the Old Spanish lexicographical effort at the University of Wisconsin-Madison as well as the parent organization under which that effort has been carried forth can hardly be classified as infants. The Wisconsin Seminary of Medieval Spanish Studies has for nearly half a century been engaged in the confection of an extensive Old Spanish dictionary, the inception of which was made possible back in 1935 through a generous grant from the American Council of Learned Societies, and work was begun on what was to ultimately become an eleven million citation slip file of the thirteenth-century texts of Alfonso X, the Wise of Spain. (See photograph.)

Following a technique which had been employed by the Early Modern English Dictionary project at the University of Michigan, 3 × 5 inch slips with the keyword of each slip underlined in red in its context were produced. When filed in alphabetical order, the file constituted a concordance to the text. In those pre-computer days this method was doubtless the most efficient way of creating such a citation file.

In 1937 the Seminary of Medieval Spanish Studies was placed under the directorship of Lloyd Kasten. In the intervening years Professor Kasten, through the good support of the Graduate School of the University of Wisconsin, continued expanding the dictionary files by means of the paper citation slip procedure until by 1970 the lexical materials on hand at the Seminary had grown to their present eleven million citation entries.

Although Professor Kasten had on numerous occasions contemplated an eventual role of the computer in the project, it wasn’t until 1971 that the two of us began experimenting with rather rudimentary machine-aided techniques with an eye toward future computerization of many aspects of the work. Closer examination of the excellent computing facilities on the Madison campus convinced us to close the paper citation files to any new vocabulary and reconsider the purposes of that dictionary. Since the manually-compiled Dictionary of Medieval Spanish had been conceived of as a defining dictionary from its very beginnings, it was decided that it would be published as a definitional manual, primarily intended to aid students and scholars in the reading of Old Spanish texts. But what of an even more exhaustive, statistically representative text citation file and lexicon of the Old Spanish tongue, one which could serve as a more rigorous research archive for both humanists and social scientists, and especially one which would meet the demands of linguists seeking a sounder basis upon which to formulate their notions regarding the chronological and dialectal development of the Spanish language of the Middle Ages? So vast an archive as we envisaged establishing could only hope for success over a reasonable number of years if the file creation and manipulation techniques adopted made heavy use of modern electronic technology.

During the year prior to the initial grant request submitted to the National Endowment for the Humanities in 1972 for financial assistance in undertaking the changeover from manual to computer-based methods, Seminary staff members began rather extensive research into the state of current computer capabilities and their cost effectiveness as these related to our proposed endeavor. Our resulting pilot proposal to NEH met with success and the machine-based Dictionary of the Old Spanish Language got under way.
As our involvement with computer techniques proceeded, it became clear that postponement of the actual development and implementation of specific computer programs for data processing was not only possible but indeed judicious, inasmuch as the most costly and time-consuming aspect of the work would still be human and not machine, that is, the preparation of accurate machine-readable versions of the Old Spanish manuscripts, a truly Gargantuan task which would necessarily precede any computer processing.

The first phase of the project on which four years have now been spent and for whose continuation additional NEH funds have been secured for the 1977-80 triennium, was designed to implement the changeover from manual to computer techniques of the Alfonsine materials. The ultimate aim of the project is to establish a data base, which will be referred to as OSA (Old Spanish Archive), in a form compatible with rapid retrieval of lexical and syntactic information, in addition to morphological data, chronology of words, individual vocabularies of a specialized sort, and geographical distribution. It should be of interest not only to students of the language and literature of medieval Spain but also to social scientists and historians, although its main function for the foreseeable future will be to serve the more immediate goal of the creation and publication of the Dictionary of the Old Spanish Language, the first edition of which is projected to appear in print some eight years hence. Obviously, it is toward the attainment of our more immediate goal that our primary efforts have been directed during these past several years.

The Dictionary of the Old Spanish Language will be a vast citational lexicon illustrating Old Spanish word usage through selected representative citations of medieval Spanish vocabulary in all its forms and functions as attested to in some 250 chosen manuscripts and incunabula produced between the years 900 and 1500. Moreover, DOSL entries will be cross-referenced to the modern Spanish Diccionario de la Real Academia Española, so that Old Spanish word usage may be accessed via the modern Spanish lexicon or directly through the old manifestations of the idiom.

As an initial stage of DOSL we have produced a machine-based core vocabulary of the truly reliable thirteenth-century texts of the royal scriptorium of Alfonso X that are preserved in their original manuscripts. They emanate from a well-controlled group of scribes and are very extensive, comprising some five million words of text. They vary in subject matter from the scientific to the legal and historical, and even offer a treatise on chess and backgammon. Furthermore, the language used in these texts is both lexically and morphologically complementary. Besides making possible the production of the complete vocabulary of Alfonso as revealed in the soundest textual sources as well as the creation of the corresponding editions of those texts, the past four years of work have enabled us to establish effective human and machine techniques for proceeding with the incorporation of textually, dialectally, and chronologically diverse Old Spanish literary works aimed at producing an adequately representative lexicon of the language.

The procedures that we developed for producing machine-readable text have been so satisfactory that we were able to transcribe the Alfonsine material directly from photostatic reproductions of the manuscripts in their entirety in somewhat less than two years. The staff time which was saved thereby was put to good use in initiating the transcription of a comparably authoritative corpus found in the works completed under the aegis of Juan Fernández de Heredia, Grand Master of the Knights of St. John of Jerusalem and adviser to the Avignon popes. The Heredian collection constitutes the largest single unified body of fourteenth-century Aragonese dialect texts, and therefore represents a major complement both chronologically and dialectally to the thirteenth-century Castilian of Alfonso X. Since a group of doctoral candidates became interested in the Aragonese materials, they have individually chosen to pursue thesis work related to those manuscripts and have donated their machine-readable versions to the dictionary effort.

With the completion of the Alfonsine transcription and a good head start on the Aragonese texts, we have laid a substantial base from which to branch out within the medieval period of the Spanish language by means of the selection and incorporation of additional, numerous surviving representatives of manuscript and incunabulum texts produced prior to 1501. Because of obvious limitations of time and future funding, the task of choosing which surviving versions of what Old Spanish works should be honored by inclusion in