# "Yet There is Method in It"; The *Cumulative Shakespeare Bibliography* — A Product of Project Planning in the Humanities

## Harrison T. Meserole and John B. Smith

"They are the books, the arts, the academes that show, contain, and nourish all the world"

*Love's Labours Lost,* IV, iii.

The *Cumulative Shakespeare Bibliography,* a two-part project begun in 1978 at The Pennsylvania State University and now residing at Texas A & M University, is a bibliographic database of Shakespeare scholarship and dramatic productions. Part one, containing records from 1958 through 1979, is nearing completion. Part two, extending the database back to the beginning of this century, is under way. When completed, the database will contain approximately 100,000 categorized, annotated, and verified citations from more than 40 countries. Results of the project will be available to scholars in any of three forms: a conventionally published, inclusive reference

work; current year cumulations published in *Shakespeare Quarterly*; and computer printouts of specific search requests offered through a custom retrieval service.

The database contains a wide variety of materials. In addition to all the relevant scholarship of this century that we can locate, it includes information on major productions, television programs, and radio broadcasts, naming directors, producers, principal actors, stage designers, and others. Also included are references to maps, films, recordings, pictures, and other works of art that bear on Shakespeare studies or productions.[1]

Scholars have long been aware that Shakespeare's plays touched upon virtually every aspect of his contemporary culture. History, politics, social customs of every level of society, religion, alchemy, and jurisprudence are just a few

1/ Shakespeare's unique position in world literature has encouraged an unceasing and burgeoning record of publications and stage presentations and, consequently, an active bibliographic industry to control and make accessible that record. Products of the industry range in size and focus from, say, a 30-item checklist of Shakespeare films for use in secondary-school classrooms; through the selective, specialized compilation of several hundred books and essays on a single play contained in the Variorum Edition of that play (for example, Mark Eccles' just-published treatment of *Measure for Measure*); through the more comprehensive 3,000-item collections published each year in *Shakespeare Quarterly* (United States) and the two Shakespeare Yearbooks (East and West Germany); to the massive cumulated Shakespeare bibliographies of which that of Gordon Ross Smith is the most recent (1963) and the largest (some 30,000 entries). Part one of our project begins where Smith left off (1958), brings the record up to date through 1979, and will contain an estimated 45,000 entries. Though Marvin Spevack used the computer to prepare his exemplary *Concordance* to Shakespeare (1973), our annotated *Cumulative Shakespeare Bibliography* will be the first comprehensive Shakespeare bibliography compiled by electronic data processing, the first to be exhaustively indexed, and the first to offer specialized custom searches of the database.

of the topics for which scholars have found Shakespeare a commentator. Consequently, the database is expected to be a valuable asset for research in many disciplines, not just in the humanities but in the social and behavioral sciences as well. The search capability will make it practical, for the first time, for scholars to explore fully a number of interdisciplinary topics. Comprehensive lists of citations will be available for Shakespearean references to, for example, art, music, philosophy, language, or women.

Among the most important new possiblities will be the opportunity to bring to bear on new theatrical productions the consolidated scholarship and descriptions of earlier productions. For the first time, directors, actors, and set designers will be able to review the history of productions of any play. This capability should make for more authentic as well as more diverse and daring productions, depending on the director's intent.

At its height project staff numbered eight, but almost 60 other individuals scattered around the world are directly involved, along with a number of institutions and agencies. Many academics hold the traditional view of humanistic research as the effort of a single scholar working alone with conventional materials. This research project is clearly quite different. While project research is not new in many fields, its application in the humanities occasionally raises new problems or elevates old ones to new levels of complexity (or consternation). In this article we describe the project, but we place details within the context of the problem-solving and planning process in which we confronted them. We hope, thereby, to address a diverse audience with widely varying computer backgrounds. The experienced computer user in a project enviroment outside the humanities may find our system and our approach interesting as an application of familiar techniques in a different discipline; the humanist with computer experience but little project experience may find the problem-solving and planning aspects useful; the reader with

little or no computer experience may find the descriptive portion informative with respect to how a computer can be used for one type of humanistic research.

While computer research in humanistic disciplines is not commonplace, neither is it unknown. Three major areas of activity are storage and retrieval of secondary descriptive information, manipulation of literary and other texts, and analysis of data within an interpretive context. The first category is quite broad, including a variety of bibliographies, indices, checklists, and catalogs. Our project is somewhat representative of this category but is rather larger in scope than most. The second group includes various dictionary, concordance, and textual collation projects, each beginning with some text(s) and producing output that is derived from or is a restructuring of the original text(s). The third category consists of projects that attempt to support interpretive arguments based on the application of some analytic model to texts or other forms of data. In literary studies, for example, analysis of style are often based on distributional patterns of various language features over one or more literary texts; by contrast, historians are increasingly turning to analyses of census and other demographic records to support their generalizations.

From this very brief overview, we can infer some of the factors that characterize computer research in the humanities and must be taken into account in the design, planning, and implementation of a project. First, most of these applications involve medium to large collections of data and require a good deal of attention to record format, file design, and access modes. Second, the end user is likely to be unfamiliar with and perhaps unsympathetic to computers. He is likely to work in a field with strong scholarly traditions regarding the format, appearance, and organization of research materials, and to expect a high degree of precision in the data. Third, the data and products are likely to be used by scholars from a number of disciplines, ranging from literary studies to cultural

anthropology; consequently, access and analysis systems must be highly flexible and must support an almost open-ended range of perspectives. Finally, the project itself will almost invariably have to link the efforts of a number of individuals, all highly trained but in widely different areas; this in turn places heavy emphasis on the communication, negotiation, and coordination skills required of the project leaders. In the remarks that follow, we try to show how these and other factors were melded in the planning and implementatin of the *Cumulative Shakespeare Bibliography.* We look first at the coordination and administrative activities, then at the system design; our discussion concludes with a description of the work flow and procedures that resulted from these activities.

## COORDINATION AND ADMINISTRATION

Our planning, both administrative and technical, began with the end products and capabilities we wished to develop. Less obvious were the criteria we should use in their design. There is a comfort of sorts that one can take in engineering projects that result in a product or other tangible object: the device either works within specified limits or it does not. For projects in the humanities, the ultimate criterion, particularly for reference projects of our sort, is the utility it will eventually provide for a body of scholars and students. While far less tangible, this criterion is no less real than whether the plane flies or the bridge stands. We soon realized that we needed an advisory committee to assist us in thinking through the larger aspects of design and in setting broad project policies. The group of eight senior scholars who agreed to work with us has also been helpful in explaining the project to others and in building the support in the scholarly community necessary to ensure the *Bibliography*'s acceptance, hence its utility.[2]

2/ The *Cumulative Shakespeare Bibliography* advisory committee:

One of the first questions we took to this group was the matter of the subject taxonomy that would determine the overall structure of the database and, to a measurable degree, the modes of access to it. Earlier bibligraphies had used extremely elaborate category systems — one took 40 pages to enumerate — or systems rendered obsolete by computer search capabilities. Our task was to provide enough structure to guide the scholar while not requiring a long period of familiarization. The advisory group was particularly helpful to us in thinking through the implications of various category schemes and in determining what information to embed in the taxonomy and what to include in the data record itself. The final result was a four-level hierarchial system as outlined by the example in Figure 1. As we show below, the framework not only determines the overall order of the *Bibliography* but also guides access to it for individual computer search requests.

Computerization of bibliographic projects offers the obvious advantage of permitting updating and correction of the file. Nonetheless, in projects such as ours it is highly desirable to make the database as complete and accurate as possible during the period of primary funding. Shakespeare is one of those few authors who has world stature. Consequently, relevant scholarship and productions appear not only in English-speaking or Western nations but throughout the world. To assist us in scanning publications not readily available in the United States, we have a committee of correspondents consisting of more than 40 volunteers. This group sends us citations to be included and, for items in unfamiliar languages, provides translations of titles, descriptive terms, and annotations.

John F. Andrews, National Endowment for the Humanities
David Bevington, University of Chicago
Maurice Charney, Rutgers University
Alan C. Dessen, University of North Carolina
Roland Mushat Frye, University of Pennsylvania
Michael Kiernan, Pennsylvania State University
Marvin Rosenberg, University of California, Berkeley
Susan Snyder, Swarthmore College

General Shakespeareana
Play Groups
Individual Plays
.
.
.

Hamlet
    Scholarship and Criticism
        Bibliographies and Checklists
        Editions and Texts
        Translations and Adaptations
        Sources and Influences
        Textual and Bibliographic Studies
        Criticism

    Production and Staging
        General
        Bibliographies and Checklists
        Reference Works
        Actors, Acting, Directing
        Dance, Ballet
        Film, Cinema, Radio, Television
        Music, Opera
        Readings, Recordings
        Stage and Theatre History
        Stage Productions
        Theatrical Techniques, Scenery, Lighting, etc.

1 Henry IV
.
.
.

Figure 1:    Example of the four-level hierarchial system im-
             plemented in the *Shakespeare Bibliography*.

Keeping track of the correspondents and our advisory committee is a time-consuming but essential administrative task for the project. Additionally, we have, or are in the process of developing, formal working relations with some six institutions or corporations. Each of these relationships has taken a great deal of planning and, in some instances, negotiation. Our primary funding source for the past three years has been the National Endowment for the Humanities. Most of the administrative as well as technical structure we discuss here was worked out prior to initial funding. Consequently, the preparation of proposals and the background work required took a number of man-months of effort. Hard-copy publication involves working agreements with four institutions. Annual current bibliographies are published in *Shakespeare Quarterly*, issued by the Folger Shakespeare Library in Washington. Our working arrangement with them includes scheduling, division of costs and responsibilities, and frequent *ad hoc* decisions. Since final copy is set by a computer-driven photocomposer, we had, in turn, to negotiate technical agreements for formatting and for the codes to be placed in the data to produce copy in the desired form. A similar set of negotiations lies ahead with the publisher of the final cumulative edition, and with his printer. While technical negotiations for photocomposer codes can be exacting and tedious, the benefits are substantial. To produce a page of camera-ready copy for the *Cumulative Bibliography* would cost about $40 for typesetting manuscripts in typescript, compared with $7 for "manuscripts" coded on magnetic tape. Thus, for the estimated 2,300 pages of the first 22–year segment (1958–1979) the saving amounts to almost $80,000. For projects in the humanities, such savings can mark the difference between a research resource that is an attractive publication prospect and one that would require a large subvention. Finally, our relationships with Pennsylvania State University and Texas A & M University have required and will continue

to require considerable attention. Both universities have been highly supportive of the project, but numerous details and arrangements have had to be worked our concerning equipment, space, personnel policies and procedures, interim funding, and access to the computer.

While most researchers would anticipate the need for planning in the technical dimensions of a computer research project, some may not anticipate the equally important and sometimes more difficult planning and problem solving involved in the coordinative and administrative aspects of the project. The tasks we faced will not transfer to all projects, but formalization of the relationship with the scholarly community, development of measures to ensure completeness, and coordination of activities with other institutions and corporations are probably tasks encountered by almost all research projects.

SYSTEM DESIGN

The broad system objectives — to support a dynamic database, to provide flexible search capabilities, and to support direct photocompositions of end products — were set as part of the overall project objectives; the detailed specifications, however, were negotiated between the Director and the Associate Director of the project.

The Director, a traditionally trained humanist, had gained a good deal of familiarity with computer systems while guiding the *Modern Language Association International Bibliography* through its initial computerization during the early 1970s; the Associate Director, while having traditional training as a humanist, also was an experienced system designer and programmer. An early review of existing bibliographic services and systems indicated that none fit very closely the needs of the project in terms of costs, access, and capability. Consequently, we decided to develop our own system, based on an earlier system designed by the Associate Director, to meet

the needs of our project precisely, efficiently, and economically. Had the necessary technical expertise not already been available to the project, our decision in this matter probably would have been different.

The process of negotiating detailed specifications and system capabilities proved to be quite interesting. Our relationship on these issues soon became apparent as one of friendly, constructive opposition. The Director negotiated from the position of trying to make the final product conform as closely as possible to traditional bibliographic expectations; the Associate Director negotiated from the position of trying to make the system as efficient and flexible as possible. The broad middle group that we shared greatly facilitated these discussions and enabled us finally to develop a system that satisfies both sets of criteria. But we each had to give a little. As examples, we wished to enter all names in the database in index order — Last, First, Middle, Jr. — to provide a consistent form for searching; on the other hand, traditional bibliographic format calls for second and subsequent names in multiple-author citations to appear in normal order. The compromise resulted in nearly two pages of closely written PL/I code to unscramble names: score one for tradition. Reviewers of play productions often sign their reviews with initials. For some reason certain Slavic-language reviewers seem prone to use lower-case initials. When the index program sorts these "names," they float to the top, above words beginning with upper-case A. Rather than modify the standard SORT package or develop a seldom-used separate SORT field, we allowed this order to stand: score one for efficiency.

The data processing system includes two groups of programs: those that perform various "housekeeping" chores in the process of developing and maintaining the database, and those that provide access to the database and produce end products. The "housekeeping" set consists of three programs. The first restructures the input from a line format suitable for

typing and proofreading to an internal variable-length record better suited for processing. The second sorts new entries and merges them with the master file. The third program produces a back-up copy of the master file and also extracts selected records from the database for editing and subsequent re-entry.

The second group of programs provides access to the database. Most access and formatting functions are provided by a single search program. This module can select items according to a search expression or reproduce the entire master file, select and rearrange various fields of information, format the content, re-sort the output on any sequence of internal fields, and produce an internally coded data set suitable for direct input into a photocomposer. The second access program produces indices to the master file from the contents of designated fields; thus, separate indices can be produced for authors and for descriptive terms or they can be merged.

As we mentioned above, our master file uses a taxonomy with four hierarchical levels. To permit possible eventual merger with other data bases and to accommodate other applications, we designed the system to support hierarchial taxonomies of any (practical) depth. Similarly, all fields within a record are variable in length, marked internally by codes rather than by position within record. The system will support many different category schema for fields; our task was to determine the most advantageous balance between over- and under-division of the record. For output, the system required minimal internal division; however, to meet a wide variety of anticipated search requests, we need to identify the context in which data appeared so that we could differentiate, say, between the books and articles written by an author and reviews of the author's work.

The record format that resulted has five types of fields. The first field is the subject taxonomic code, which determines the segment of the master file in which the individual record or citation will be placed. This field is followed by a half-dozen

or so fields where the factual data of the citation are stored: author, title (separately identified as a book or an article or chapter contained in a larger publication), journal or publishing information, and so forth. In most citations there is a third type of field containing a narrative annotation of two or more sentences. Fourth, we include, in separate fields, names for the specialized indices we generate: names associated with the theatre (actors, producers, directors) or names mentioned in the work referred to (other authors, philosophers, historical figures). Finally, we assign to each citation descriptive terms drawn from a controlled thesaurus of approximately 800 terms. These terms are indexed separately and can be used in custom searches. Figure 2 illustrates a bibliographical record as edited and ready for input, as entered by field for its primary listing, and as presented in the final sorted and classified printout.

Full Boolean logic is available for specifying search expressions involving content as well as form and subject taxonomies.[3] For example, all items categorized within the subject taxonomy under *Hamlet* can be extracted; or all items that contain the word "Hamlet" anywhere within them; or all items that contain the word "Hamlet" within the title of

---

3/ Readers unfamiliar with Boolean logic may regard it as a precise, but limited, syntax than can be used to express relations among search keys. Our retrieval program uses Boolean expressions to seek out and display all the records (and only those records) that meet the condition(s) specified. A search key may be any designated portion of the bibliographic record: the subject-taxonomy code, any word or string of words, or any form-taxonomy code. Boolean logic allows the researcher to specify contingent relations, such as two keys both of which appear in a record (a logical AND relation); either of two keys appearing in a record (a logical OR relation); or a key not appearing (a logical NOT relation). The search expression can contain multiple keys joined in complex relational structures:

a book; or any logical combination of these types of keys.[4]
These search capabilities, permit a great deal of flexibility for
access and have encouraged a number of other projects, includ-
ing many that are not bibliographical, to adopt our software.

WORK FLOW

Having traced the planning and problem-solving steps in-
volved in the design of the system and the organization of the
project, we now briefly describe the procedures and work
flow that have resulted. Bibliographic records came from many
sources, but the vast majority are derived from three: journals
and books scanned by our internal staff; titles supplied by
our correspondents in the United States and other countries;
and bibliographies and checklists published previously.
Whenever possible these records are all traced back to the
original sources for verification by our staff or by a correspon-
dent. While this procedure is time-consuming, we have found
that nearly a third of the records require some type of correc-
tion. Verified items are then coded within the subject
taxonomy, the annotation added and edited for style, and the
descriptive terms from the controlled thesaurus appended.

---

for example (%30.14 or Hamlet or Ham.) and not %.20. In our system,
%30.14 is the subject-taxonomic code for items principally concerned with
the play *Hamlet*, while %.20. is the form-taxonomic code for separately
published works, typically books. Thus, this retrieval would locate all items
that are classifed under the *Hamlet* heading, or that contain the word
"Hamlet", or that contain the internationally recognized abbreviation
"Ham.," but (and) which are not books — that is, all essays, chapters,
reviews, productions, recordings, etc. that deal with *Hamlet*.
4/ We also use the internationally accepted set of official abbreviations for
Shakespeare's works (for example, *Ham.* for *Hamlet, Tro.* for *Troilus and
Cressida, PP* for *Passionate Pilgrim*, and we can apply search strategies
using these sigla as keys.

A

1&2H4/Schol/Crit
%30.20.05.30

Gottfried, Martin.
"Shakespeare's Nuisance." *SatR*, 23 June 1979, p. 30. [On the character of Falstaff, not only in *1&2H4*, *H5*, and *Wiv*. but also in Verdi's opera et al.]

%.63 Verdi, Giuseppe
%.64 honor, parody, fool
CRS        *Wiv*/Schol/Crit %30.50.05.30
           *H5*/Schol/Crit %30.22.05.30

B

%30.20.05.30
%.10 Gottfried, Martin.
%.25 "Shakespeare's Nuisance."
%30 *SatR*, 23 June 1979, p. 30.
%.35 1979
%.40 [On the character of Falstaff, not only in *1&2H4*, *H5*, and *Wiv*. but also in Verdi's opera et al.]
%.63 Verdi, Giuseppe
%.64 honor, parody, fool
%.80 6828
%end

C

1597 Coffman, Barbara Jean. "Acting That Argument." s.v. R2/Scholarship/Criticism

1598 Gottfried, Martin. "Shakespeare's Nuisance." *SatR*, 23 June 1979, p. 30. [On the character of Falstaff, not only in *1&2H4*, *H5*, and *Wiv*. but also in Verdi's opera et al.]

1599 Hardy, Barbara. *Dramatic Quickleyisms: Malapropric Wordplay Technique in Shakespeare's Henriad*. 2 vols. (Eliz. & Ren. Studies 85.) Salzburg: Inst. für Englische Sprache und Literatur, Univ. Salzburg, 1979. 530 pp. [Comic lang. in the subplot mirrors the main plot in *1&2H4* and *H5*.]

Figure 2:     A bibliographical record (A) as edited and ready to be entered into the data base, (B) as entered by field for its primary listing, and (C) as presented in the final sorted and classified printout (item 1598).

Once edited and complete, items are ready to be entered into the database.

In our environment, typing and correcting items on line is more cost-effective than off line. Consequently, we use Penn State's general editor and file system, INTERACT. The items are proofread once in a line format, one field per line, and then in a bibliographical format. Corrected files are duplicated, then sorted and merged with a submaster file for each year of publication. Once a year, the current year's cumulation is processed and used to produce the annual "Annotated World Shakespeare Bibliography" published in *Shakespeare Quarterly*. Current and retrospective cumulations are periodically merged into the master file. The infrequent item found to contain an error is extracted and reinserted correctly with the next batch of entries.

The master file is available for searching at any time. To date, we have provided individual searches on an occasional and *ad hoc* basis. Later this year we expect to announce a custom retrieval service, which will mark a major new dimension for our project. Our goal is to provide a wide range of services at a reasonable cost. While the operation will be a non-profit one, we hope to develop a level of service and a pricing policy that can support not just the continuation of this service but future updates to the *Bibliography*, once the retrospective cumulation is complete and major financial support is expended. This may well be the most challenging planning and problem-solving task we will confront.

Our final products will include two library editions of the *Cumulative Bibliography*: one for the 22–year period 1958–1979, the other for the period 1900–1957. Each will be indexed separately, and they will also be indexed collectively. Both will, of course, be set by photocomposer from the master file. If our plan for financial self-support succeeds, we expect to publish updates at five-year intervals beginning in 1985.

CONCLUSION

We would not wish to leave the reader with the impression that we anticipated all the problems we encountered, or that we have always made the right decisions. Yet our project has been of sufficient size, complexity, and sensitivity within the primary discipline areas that we sought to anticipate as many factors as possible and include them in our design. One of the most important things we have learned about project design is that clear decisions often lie at the end of winding and littered paths. Any computer system used in such a project will reside not just in the hardware but within a working context, an intellectual or scholarly tradition, an organization, and a network of agreements with other individuals and institutions. A computer system will work smoothly only if all these other systems functions smoothly, as well.