Research Article

Enhancing Archival Description for Public Computer Conferences of Historical Value: An Exploratory Study

ANNE J. GILLILAND-SWETLAND AND CAROL HUGHES

Abstract: This paper reports on a pilot study that explored new approaches to the description of computer conferences. The authors tested methodologies for making archival description for public computer conferences of historical value more rigorous and accurate than would be possible using only traditional archival approaches. In their study of the Wing:Span public conference at the University of Michigan, the authors found that a considerable amount of additional and more precise descriptive information could be generated by using unobtrusive observation and statistical techniques to gather and analyze data. They recommend replication of the most productive and cost-beneficial of these methods in the study of other public conferences to assess their potential value as tools to enhance existing archival descriptive methodology.

About the authors: *Anne Gilliland-Swetland is an associate archivist at the Bentley Historical Library, where she is working under a grant from the National Historical Publications and Records Commission to investigate the potential archival value of computer conferences. She worked as archivist and records specialist at the University of Cincinnati from 1985 to 1990 and as project assistant for the American Library Association (ALA) Archives at the University of Illinois from 1983 to 1985. She is currently a full-time doctoral student in the University of Michigan School of Information and Library Studies, where she is investigating the archival management of electronic records in academic environments.*

Carol Hughes is a lecturer and full-time student in the Ph.D. program at the University of Michigan School of Information and Library Studies. In 1992, she was coauthor, with Richard M. Dougherty, of "Library Cooperation: A Historical Perspective and a Vision for the Future," the cowinner of the 1992 Blackwell North America Scholarship Award given by the ALA's Association for Library Collections & Technical Services. Her other recent publications include "A Comparison of Perceptions of Campus Priorities: The 'Logical' Library in an Organized Anarchy," which appeared in the July 1992 issue of Journal of Academic Librarianship, *and* Preferred Futures for Libraries, *coauthored with Richard Dougherty and published by the Research Libraries Group in 1991.*

The authors wish to thank Joseph Janes of the University of Michigan School of Information and Library Studies for his assistance with the statistical analysis contained in this paper.

TRADITIONAL ARCHIVAL TECHNIQUES FOR the description of collections involve three processes:

- Using the archivist's knowledge about the medium in which the materials were created and the functions that the records or documents served.
- Understanding the arrangement and structure of the records.
- Identifying the record creators and major topics.

Relying solely on this approach in the description of electronic computer conferences, however, is problematic. Archivists know less about the origin and contents of such conferences and understand less about the nature and function of records in an electronic medium than they do about those in paper media.

Computer conferencing was originally conceived of as an electronic forum that could be substituted for face-to-face meetings and used for group decision making and conflict resolution. Over the past twenty years, particularly in the academic environment, this role has expanded considerably and now encompasses both task-oriented and social functions. Archivists familiar with using Bitnet and computer network "listservers" would recognize a computer conference as a similar but much more highly structured communication medium, which can be used for the online discussion of multiple issues related by a common theme or function. Thus, a conference might exist on a university campus as a forum where students, faculty, and staff could discuss health issues. Individual issues—such as health insurance options, university healthcare providers, abortion, and Acquired Immunodeficiency Syndrome—would each be structured as a distinct "item," within which discussion of that particular subject would take place. Conferences are usually created asynchronously and are linked vertically and horizontally by pointers to "prime" items (earlier discussions that were the source of the new topic) and related "items."[1] These linkages allow participants to read or scan items and their associated responses either chronologically or by subject-relatedness.[2] This paper describes a research project that was designed to explore new and more rigorous data-gathering methods to complement traditional approaches for the archival description of computer conferences of historical value.

Background

The Bentley Historical Library has been conducting research since early 1992 under a grant from the National Historical Publications and Records Commission (NHPRC)[3] to determine whether public conferences at the University of Michigan exhibit archival value for documenting the intellectual, cultural, and social life of a major academic institution.[4] A complete

[1]A "prime" is a preceding item, the discussion of which prompts a participant to initiate a new item on a related topic. "Relators," or descriptive phrases cross-referencing topically related items, are assigned within ongoing items by the conference organizer.

[2]Ellen M. Pearson provides a succinct description of the process as the creation of "an ongoing database of all text contributed by the conference members. Members may search for and retrieve stored text at any time. Typically, participation in the conference is asynchronous. . . . Each participant sees all the others' statements and may comment on those already entered and/or add new thoughts to the discussions. The conferencing system software tracks all entries, linking statements, and comments thereto, so that members may read or proceed through the messages either chronologically or logically. The software also tracks each member's individual online session so that when he next joins or signs on, he is notified of the numbers of new or unread messages." See Ellen M. Pearson, "Computer Conferencing for Enhanced Communication: Its Potential for Academic and Research Communities," in *International Library Cooperation: 10th Anniversary Essen Symposium, 19 October–22 October, 1987,* ed. Ahmed H. Helal and Joachim W. Weiss (Essen: Universitatsbibliothek Essen, 1988), 328–37.

[3]NHPRC Grant No. 91-113.

[4]Some of the archivist's primary professional responsibilities are to preserve and transmit cultural heritage. College and university archivists perform these responsibilities by documenting what Helen Samuels

report of all findings of the NHPRC project will be published when the project is completed. The project described in this paper was a small pilot study, distinct from the main thrust of the project, conducted to explore new approaches to the description of computer conferences.

Computer conferencing using Confer II software was started at the University of Michigan in 1975. Since 1981, more than 3,100 conferences have been hosted, with over 165,000 membership registrations. In early 1992, log-ins to conferences were running at about one every thirty seconds.[5] Depending on the nature and function of the conference, membership may be private or public. There are many types of private computer conferences, including administrative, course-related, and social. Public conferences are those in which participants are not restricted to university affiliation or course membership status, and membership is granted on request. All entries can be read by all conference members and by nonmembers who choose to observe or audit the public conference.

The University of Michigan is not unique in its use of computer conferencing. Many other universities, including Oakland University, Pennsylvania State University, the University of California at Berkeley, the University of Guelph, the University of Maryland, and Wayne State University have increased their use of computer conferences for a variety of activities. Some of these sites also use Confer II software, a package that well represents the typical characteristics of conference software. Conferencing within the academic environment can be used for collaborative research, coauthorship of manuscripts, class instruction, professional socialization, or "intracommunity networking," the linkage between academic groups on a campus for the discussion of topics of mutual interest.

The Bentley project archivists have determined that many conferences they have observed as part of the archival appraisal process[6] have substantial value for documenting the academic environment and its various constituencies and subcultures in a way not reflected by the administrative record. They anticipate that accessioned conference material will interest not only historians but also sociologists, anthropologists, and educators. The development of accessioning and storage mechanisms is now under way. The next logical step is developing for each "archived" conference finding aids that reveal to potential researchers as much as possible about each

has defined as the seven basic functions of academic institutions: (a) conferring credentials, (b) socializing, (c) conveying knowledge, (d) maintaining culture, (e) advancing knowledge, (f) providing service, (g) sustaining itself. Academic archivists are beginning to examine computer conferencing, a phenomenon that has spread widely in its application throughout academia since the early 1970s, as a potentially rich documentary source for these functions of colleges and universities. See Helen W. Samuels, "North American Archival Identity," in *International Council on Archives, Second International Conference, Ann Arbor, MI, 9–13 May 1989, Proceedings,* ed. Judith A. Koucky (Washington, D.C.: National Archives and Records Administration, 1990), 83–86. See also Nicholas C. Burckel, "The Expanding Role of a College or University Archives," *Midwestern Archivist* 1 (1976): 5. Burckel writes that "another collecting focus [of college and university archives] should be the intellectual and cultural atmosphere which the university engenders. This can hardly be determined from a look at transcripts or college catalogs."

[5]Electronic mail communication from Robert Parnes, developer of Confer II software, 18 October 1991.

[6]By observing over fifty-five active conferences online, appraising them for historical value, and assigning subject descriptors to indicate content, the project archivists have been able to compare date and subject coverage of the conferences with collections in traditional formats held by the Bentley Historical Library. Conferences thus far appraised have contained documentation of aspects of campus life from perspectives and a currency not evidenced by existing Bentley collections. Approximately 75 percent of the appraised conferences have substantial archival value. (Although until the comparison with traditional materials is completed, the amount of content overlap between the two types of materials, and thus the percentage of conferences that actually will be accessioned, cannot be definitively stated.)

conference. Particularly important questions appear to be how the conference was used, when it was active (e.g., what for, and when?), by whom (e.g., how widespread was the use, and were certain participants key figures or "dominators"?), and which topics were discussed.

As stated earlier, traditional archival techniques for developing such descriptions involve using the archivist's knowledge about the medium and the functions that caused records to be created, in addition to examining the records' arrangement and structure and noting significant players and topics. There are several reasons why this method should be enhanced for use with electronic conferences. First, applying such an approach to screens of data is more difficult than applying it to paper documents, where the physical extent and chronological coverage are more readily apparent. Second, Bentley archivists are interested in making available online to researchers at remote sites not only U.S. MARC records of computer conference holdings but also actual inventories to such electronic materials. Ideally, researchers should be able to use these inventories and the narrative descriptions contained in their scope and content notes to determine which electronically "archived" materials they wish to retrieve and view. They should be able to conduct all these processes online from their own computer terminals, at minimal expense in terms of computing time and without the mediation required for access to traditional materials. For these reasons, the narrative descriptions contained in finding aids to electronic materials will have to be much more self-explanatory than those currently created for more traditional materials.

Third, archivists' observations about participant dominance in any one conference might be subject to bias if they observed that the same participants were active in other conferences. This bias might cause archivists to overestimate the role assumed by certain people and, as a result, to incorporate incorrect or inaccurate assumptions into the conference descriptions.

These reasons caused us to consider whether conferences could be more effectively analyzed and partly described using automated techniques. The fact that the archival materials in question are, for the first time, in digital form also led us to look for ways to use not only new "built-in" access points provided by the conference software itself but also new methods and statistical techniques to enhance traditional descriptive approaches.

Margaret Hedstrom, a pioneer in the field of electronic archival records, noted in a 1991 presentation that "some of the difficulty we have had generalizing about or building on the results of electronic records projects is attributable to the lack of sufficiently rigorous methodologies. Most of our research and program development efforts have been built on single case studies, and we are uncertain how to apply what we learn from one case to the next." This call for rigor was echoed at the same meeting by a presentation by Tora Bikson of the RAND Corporation.[7] To test whether traditional archival description *could* be enhanced in a more rigorous and replicable way, we devised a research project to explore in detail the use patterns and structure of one public computer conference, using observational data taken unobtrusively from the active files of the online conference itself. The fundamental assumption we tested was that there is some transfer of the applicability of traditional archival descriptive paradigms into the new medium of computer conferences, and that what is needed is not a new descriptive methodol-

[7]Margaret Hedstrom, "Understanding Electronic Incunabula: A Framework for Research on Electronic Records," and Tora K. Bikson, "Research on Electronic Information Environments: Prospects and Problems," papers presented 24 January 1991 at the Working Meeting on Research Issues in Electronic Records, Washington, D.C.

ogy but an enhanced version of traditional archival description.

For the purposes of this exploratory study, we decided to concentrate on only one mature, public conference that the Bentley staff has determined to be of considerable archival value. The conference we chose was Wing:Span, which is dedicated to the discussion of women's issues. Wing:Span has the potential for yielding important material for the study of the experiences and concerns of women in academia. In making this decision, we acknowledge and expect that other types of conferences might exhibit very different activity patterns. But the most useful of the methods explored here might be applicable to them as well, as a means of identifying and characterizing those patterns.

Literature Review

The first step in this study was conducting a literature review to see what work had already been done by archivists and social scientists on computer conferences. No research literature exists for the archival profession on this form of electronic communication or on the application of statistical or other analytical techniques in archival descriptive processes of computer conference records. Trudy Peterson, among others, has argued on a more general basis, however, that, "basic archival principles apply to records created by the new technology, although in most instances the principles have to be amplified somewhat." She anticipates that description will be the "key to records in the new technological formats."[8] A considerable amount has been written regarding computer conferencing in psychology and the social sciences relating to content analysis and the quality of interaction among participants, group decision making, communications norms, and computer-mediated interactive educational processes.[9] However, although

[8]Trudy Huskamp Peterson, "Archival Principles and Records of the New Technology," *American Archivist* 47 (Fall 1984): 383–93.

[9]For example, T. Andrew Finn has called for an increase in research outside the laboratory on conferences concerning unstructured tasks to provide a better understanding of participant interaction patterns. See T. Andrew Finn, "Process and Structure in Computer-Mediated Group Communication," in *Information and Behavior,* vol. 2, ed. Brent D. Rubin (New Brunswick: Transaction Books, 1988), 167–93. Ronald Rice, Jane Siegel et al., and James W. Chesebro give overviews of research on computer-mediated interaction and the quality of the content created by this group process when compared with face-to-face or written communication. See Ronald E. Rice, "Computer Conferencing" in *Progress in Communication Sciences,* vol. 2, ed. Brenda Dervin and Melvin J. Voight (Norwood, N.J.: Ablex, 1980), 216–40; Jane Siegel, Vitaly Dubrowsky, Sara Kiesler, and Timothy McGuire, "Group Processes in Computer-Mediated Communication," in *Organizational Behavior and Human Decision Processes* 37 (1986): 157–87; and James W. Chesebro, "Computer-Mediated Interpersonal Communication," in *Information and Behavior,* vol. 1, 202–22. See also Judith Weedman, "Task and Non-Task Functions of a Computer Conference Used in Professional Education: A Measure of Flexibility," *International Journal of Man-Machine Studies* 34 (1991): 303–18.

Rosenbaum and Snyder explore Bitnet conferences (actually listservers) "in an attempt to provide empirical evidence of the emergence of social norms in computer conferencing." See Howard Rosenbaum and Herbert Snyder, "An Investigation of Emerging Norms in Computer Mediated Communication: An Empirical Study of Computer Conferencing," in *Proceedings of the 54th ASIS Annual Meeting, Washington D.C., October 27–31, 1991,* ed. Jose-Marie Griffiths (Learned Information, N.J.: American Society for Information Science, 1991), 15–23.

Several authors have addressed the use of electronic computer conferencing in distance education, and Robin Mason, in particular, discusses the role of the conference organizer. See Robin Mason, "Moderating Educational Computer Conferencing," *Distance Education Online Symposium (DEOS) News* 1 (1991). See also Terje Rasmussen, Joergen Bang, and Knut Lundby, "When Academia Goes Online: A Social Experiment with Electronic Conferencing for the Nordic Media Research Community," *DEOS News* 1 (1991). Both articles have extensive bibliographies relating to academic conferencing. Ellen Pearson and Mary Joan Tooey discuss enhancing communication in academic and research communities through the use of computer conferencing. See Pearson, "Computer Conferencing for Enhanced Communication," 328–37, and Mary Joan Tooey, "Computer Conferencing: A Campus Meets Online," *Online* 13 (July 1989): 54, 57–60.

this extensive literature validates or legitimatizes the research area to a certain extent, it has little to say to archivists about methods for describing conferences for the purposes of historical documentation.

Methodology

Using what archivists at Bentley had noted from appraising conferences for their historical value, we then formulated some research hypotheses about conferencing activities that could be studied to test methods for enhancing archival description of a computer conference. These hypotheses follow:

1. Certain individuals would dominate the conference.
2. In terms of the number of participants, their responses, and items responded to, the actual breadth of participation in the conference would be narrow.
3. Those participants who dominated responses would also dominate the initiation of items. Therefore an analysis of the system-generated item description summaries that indicate initiators might be a quicker way for archivists to assess the dynamics of the conference than an analysis of the whole conference.
4. There would be considerable variation in the activity of the conference at different periods in the academic calendar.
5. Individual items could be classified by type, and each type would exhibit the typical life cycle found for analogous types of traditional records (e.g., ongoing, administrative activities, issues of topical interest).[10]

[10]A life cycle is "the natural cycle of usage of a set of documents from the point of creation, usually the period of highest use, to final disposition. Documents are evaluated to establish their potential for historical research during their active life." See Frank B. Evans, Donald F. Harrison, Edwin A. Thompson, and William L. Rofes, "A Basic Glossary for Archivists, Manuscript Curators, and Records Managers," *American Archivist* 37, no. 3 (1974): 415–33. We considered one item to be analogous to a "set of documents," and we were particularly interested in establishing whether one or several periods of "highest use" occurred. This would indicate that the way these materials are created and referenced differs from that of traditional paper materials.

Data Collection

Wing:Span, a public conference relating to women's issues, was selected for analysis because it is a conference that has been in existence for several years. It is likely to exhibit mature communication patterns, has a wide and very active participation representing most levels of the university community, and covers a topic that is underdocumented by existing archival collections.

The method selected for data collection was to examine what information was available from the system-generated item descriptor lists (see figure 1). Through unobtrusive observation of the conference "in vivo," we then gathered further data by manually coding and tallying the number of discussion items and participation levels in the related discussion responses. (Confer software cannot automatically generate this detailed level of data analysis.) We analyzed the data graphically, using exploratory data analysis, which is used when a researcher is not sure what to expect from the data. It includes several simple ways of graphing results so that the data reveal preliminary patterns that can then be further investigated. Where it appeared warranted from the graphs, further analysis using statistical tests was performed.

The analysis was exploratory because the underlying distribution of data was not known. But we suspected that, like analyses resulting from citation analysis studies, it would be highly skewed. For ease of comparison, we selected three six-week periods representing different points in the

Figure 1. Item Descriptor Summary

A. SNAPSHOT OF ITEM DESCRIPTOR LIST

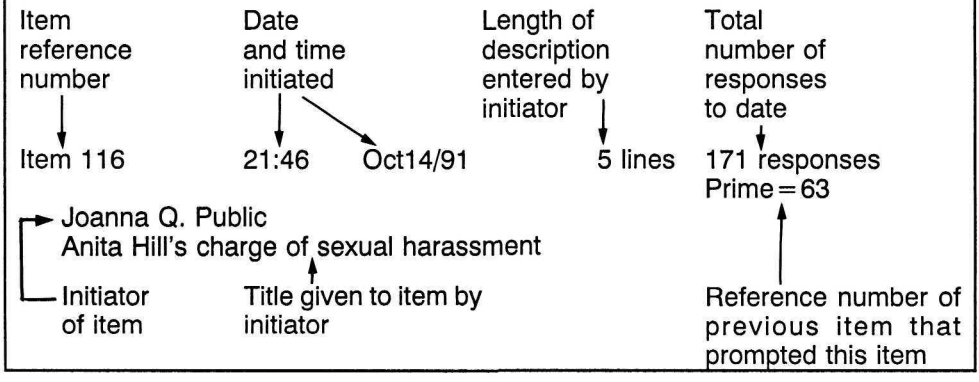
Item 13 17:18 Jan13/92 2 lines 66 responses
Name 1
The Men's Movement

Item 14 16:18 Jan13/92 5 lines 53 responses
Name 2 Prime = 13
Women's Spirituality: Awaken, lovely goddess that you are

Item 15 10:35 Jan14/92 6 lines 46 responses
Name 3
Women and TV

Item 16 11:48 Jan14/92 3 lines 27 responses
Pseud Prime = 15
Books and the role they play in defining gender roles

B. EXPLANATION OF ITEM DESCRIPTOR FORMAT



1991 academic year—23 February through 6 April, 1 June through 13 July, and 5 September through 17 October—and we counted all items in existence during any or all of those periods.[11] We reasoned that six weeks would be enough time to see the life cycle of items begin to emerge and would provide ample time for observing and comparing participant patterns.

We developed decision rules for data that might present problems for coding (such as

[11]The choice of periods that fell at the beginning and end of the two semesters, as well as during the summer, was a deliberate attempt to gather data illustrative of conference activity throughout the academic year. Having previously observed the subject content of the conferences, we were aware that several items contained lengthy discussions of two major national issues: the Gulf War, and the Clarence Thomas–Anita Hill controversy. We could have looked at these items, almost in their entirety, by slightly altering the observation periods. (As it was, the observation periods covered some, but not all, of these extensive discussions.) We believed, however, that by altering the observation periods, we would have been both allowing our own interests in these topics to possibly skew the data and doing the work of the final historical researcher, rather than that of the archivist.

the use of pseudonyms),[12] and we divided the data collection activity between us. We did code some items together, however, to check for intercoder reliability, which proved to be high; the simple counting approach left little room for interpretive error. Counts were made of the number of each participant's responses, topical items responded to, and items initiated (see figure 2). Based on the item description entered by the initiator, each item was categorized as an issue, ongoing, one-shot, or retired/deleted, and the ratio of their proportions, one to another, was determined to see if different types of items exhibited different activity patterns. "Issues" represented discussions of individual topics related to the theme of the conference; "ongoing" represented administrative items, such as conference participant introductions and games; and "one-shot" represented items, such as announcements or polls, that had a defined lifespan. The other two categories, "retired" and "deleted," represented items that were removed totally from the conference or were set aside from regular display by the organizer or the initiator.

Participant responses to every item in each period were then counted and totaled (see figure 3). Also counted were the number of items initiated each month and their primes, data easily available from the item descriptor summaries. These summaries also provided total numbers of responses per item, allowing us quickly to spot key topics without further analysis.

Data Analysis

Where applicable, more than one analytical approach was used for each hypothesis tested. First, frequency counts of the total responses in each period were graphed. The graph indicated that there might be a difference in response rates among the time periods, and we therefore tested this, using a Friedman Analysis of Variance (AOV). This statistical test looks for differences among groups when the samples (in this case, the three time periods) involve the same people measured repeatedly over time. It tests whether the differences that occur are greater than those that might occur randomly, within the 95 percent probability range. The AOV indicated that there was no statistical difference among the time periods at this level of probability. Participation in Wing:Span did not drop off significantly during the summer months, as might happen in a public conference dominated by undergraduates who leave campus for those months. A likely explanation is that most participants of Wing:Span are graduate students, faculty, and staff.

To explore further patterns of participation over time, we used information contained in the item descriptor lists to count the number of items initiated by month. This count did not reveal any unusual pattern in the number of items initiated during the entire life-span of the current volume of Wing:Span[13] (see table 1).

We then created a scatterplot to study participation patterns and, in particular, dominant individuals. A scatterplot is a graphical plotting of all the pairs of points representing the relationship between two

[12]Organizers urge conference users not to use pseudonyms, but they are used to a greater or lesser extent on most conferences. Usage tends to be either to raise sensitive or personal issues or for game or nonserious discussion items. Conferencing etiquette is that participants use only one pseudonym, which should not be someone else's pseudonym or the name of a prominent personage, but there is no guarantee even that participants using "real" names are who they say they are. We therefore decided that each name used for a response or to initiate an item would be treated as a separate participant. The actual number of obvious attempts to conceal identity during the observations was 51 out of a total 232 participants we observed. This proportion could be higher or lower, depending on the nature of the conference observed.

[13]Organizers of public conferences periodically "archive" them onto magnetic tape and start new volumes, either because their size has exceeded the limits of the software or because of some other need to make a fresh start, such as a policy change.

Figure 2. Sample Coding Forms for Manually Gathered Data

A. ITEM DATA SHEET

Item no. _____ Life-span dates: _____
Dates observed: _____
Item title: _____
Class: Ongoing ____ One-shot ____ Issues ____ Retired ____ Deleted ____
Initiated by: _____

Response no.	Date posted	Respondent

B. PARTICIPANT DATA TALLY SHEET

Participant name: _____

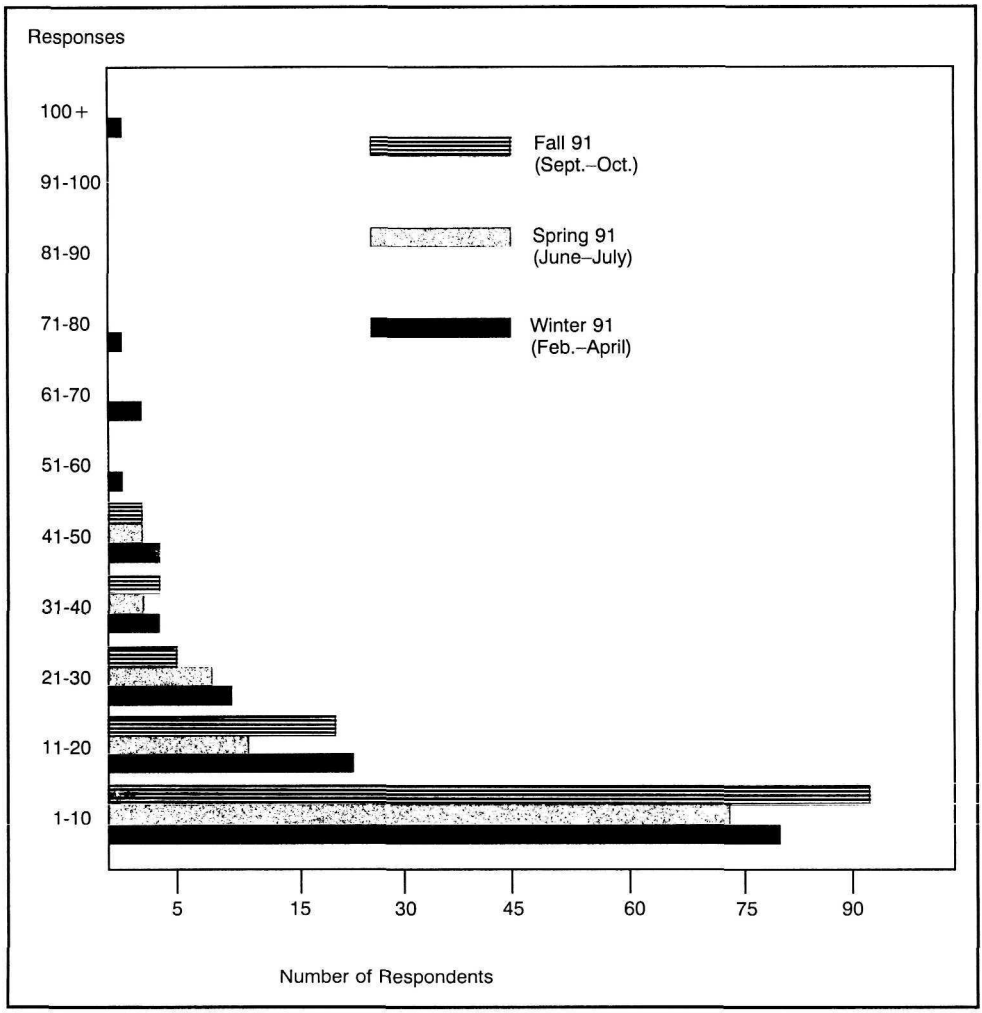
Item no.	Observation 1 response tally	Observation 2 response tally	Observation 3 response tally

C. INITIATOR DATA TALLY SHEET

Initiator name	Item nos. initiated

variables. In this case, for each person, the total number of items responded to were plotted against the total number of responses made by that person. The items are represented along the X-axis; the responses along the Y-axis (see figure 4). Scatterplots are an "exploratory" data analysis technique. They graphically display how these data points fall. A researcher may then examine the shape and direction of the overall set of points to see if any patterns indicate a relationship between the two variables. A strong correlation between two variables is indicated when points fall in a pattern close

Figure 3. Wing:Span Participation



to a straight line. In this case, the scatterplot indicated a strong correlation, and we therefore proceeded to the next step, applying the most common statistical test of correlation, the Pearson *r*.

Correlation is a statistical test for the strength of a relationship between two variables. A value of *r* close to +1.0 or −1.0 indicates that there is a perfect linear relationship between the two variables. As one basic statistics text expresses it, "*r* measures the degree to which a straight line relating *X* and *Y* can summarize the trend in a scatterplot."[14] The analysis of the relationship between the number of items re-

[14]Janet T. Spence et al., *Elementary Statistics,* 5th ed. (Englewood Cliffs, N.J.: Prentice-Hall, 1990), 203. The correlation, $r = 0.902$, is spuriously high. The number of responses made obviously is somewhat dependent on the number of items responded to. Because the two variables are not totally independent, the value of *r* we obtained was somewhat inflated. Because we wanted to know whether it was possible to predict the number of responses made within the conference, calculating the correlation was appropriate because there is no way to measure responses independent of items responded to.

Table 1

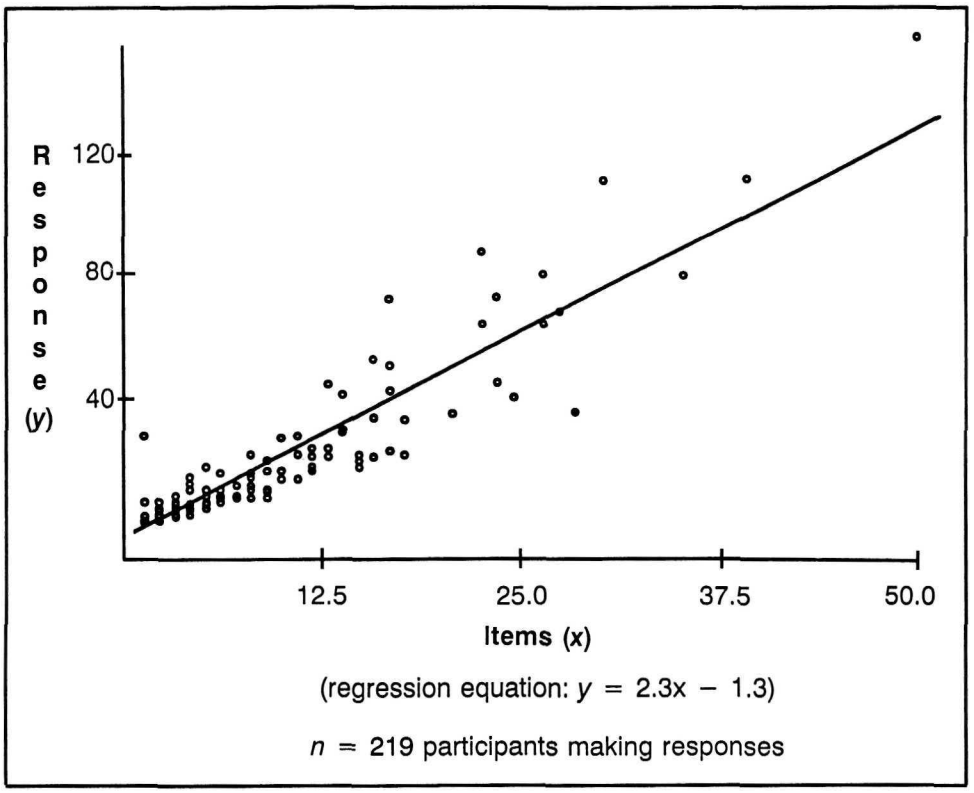
Number of Items Initiated Each Month in Current Wing:Span Volume	
October 1990	14
November	8
December	14
January 1991	5
February	8
March	14
April	18
May	6
June	7
July	5
August	7
September	9
October	5
November	9

sponded to and the number of responses made yielded an r of 0.902. This extremely strong, positive relationship appears to indicate that as items responded to increased, responses made would also increase at a *predictable* rate. R^2, a measure that in this case represents the degree to which the overall variation in items responded to accounts for variation in the number of responses made, indicated that over 80 percent of the variance between the two variables was accounted for; only 20 percent of the variance might have come from other, unknown, variables. Taking what appeared to be a very strong linear relationship, we proceeded to calculate a regression line from the data points. A regression line is a prediction measure; in this case, it can be used to predict the number of responses made from any given number of items responded to. The regression line that fits most of the data points quite closely and that appears to have good predictive ability is shown in figure 4.

It is possible to use the graph to spot dominators and outliers very clearly and to isolate them for further analysis. Dominators are participants who are very active in the conference in terms of initiating many items and making many responses. Outliers are points representing participants with participation patterns that are strikingly different from the trend shown by the graph. The outlier close to the $x=y$ line (2,31) represents an individual who responded many times to only one or two items. The most striking outlier (52,156) represents the organizer of the conference. This graph clearly shows how she dominates the conference in terms of both her total number of responses and the number of items she responded to. Other points that lie high in the scatterplot also represent dominant participants. These points, however, represent only about 10 percent of the total number of participants; the remaining 90 percent are represented by the cluster of points falling beneath the x coordinate of 20 and the y coordinate of 30. Looking at this spread, one can also begin to hypothesize about other characteristics of the conference, based upon the size of r and the slope of the regression line.

Another aspect of participation that interested us was the relationship between those who initiated items and those who responded. We had speculated that many of these would be the same people (that is, those who "dominate" the conference). There were a large number of respondents indicating very broad participation, thus our hypothesis that breadth of participation would be low was not supported. However, we discovered that only 21 percent of those responding to items initiated 100 percent of the items initiated by respondents. These figures indicate that 79 percent of the respondents never initiated any items. Most surprising, however, was that 5 percent of the total number of participants *initiated* 14 percent of the items but never *responded* to any items. This ratio would probably vary considerably from conference to conference. In this instance, the titles of the items

Figure 4: Correlation of Items and Responses



initiated by persons who were not respondents seemed to indicate that they were using these items to start discussions on provocative or personal issues (help-seeking behavior).

We looked at the proportion of items in each of the five categories (ongoing, issues, one-shot, retired, and deleted) to see which category predominated, and we found a considerable difference in frequency among the categories. The overwhelming majority of items were issues. In total, 116 items were categorized (which proved to be all items in existence from the start of the volume until 17 October 1991). The breakdown by category was 91 (78%) issues, 17 (15%) ongoing, 4 (3%) one-shot, 2 (2%) retired, and 1 (1%) deleted. This particular breakdown reflects something of the character of the conference, and it would be expected to vary significantly with different conferences.

The depth of structure within the conference was examined by counting the number of primes and relators, as well as by using the online summaries that were created by the item initiators and the organizer. The total for each of these "linking" structures was 17 primes, 29 relators, and 2 summaries. Within the framework of 116 items, this distribution represents a relatively low level of internal subject-related structure. This could mean that the issues being discussed were only very loosely re-

lated by the general theme of the conference, or it could indicate the lack of significance the participants attached to constructing subject-relatedness within the conference.

Close inspection of patterns of activity within items provided very useful data. By "eyeballing" the number of responses per participant in each observation period, we were able to identify the points at which some participants joined and left the conference, a detail the software does not track. We also examined the total number of responses to items in each observation period, and we were able to see the rise and fall (and sometimes another rise) of activity in a particular item. This gave some indication of the life cycle/activity patterns of different types of items. Some items had unusual patterns of repeated activity and inactivity, which is something archivists do not see in traditional records. We could not have obtained this detail on activity patterns from item description summaries alone.

Findings About Wing:Span That Could Be Used for Its Archival Description

As a result of this study, we discovered many aspects of Wing:Span's activity and nature that we would not have been able to discern intuitively or with the help of the summary item descriptions. These include the following:

- There is low use of the software's internal structural mechanisms, but the organizer dominates discussions and initiation of items.
- Some difference in usage occurs at different times in the academic calendar, but the conference is still quite active during the summer, possibly indicating a hard core of users who are either staff or graduate students.
- The Wing:Span conference deals primarily with topical issues (i.e., issues relating to women) rather than with ongoing administration or games.
- Few items exhibited life cycles similar to traditional archival materials.
- Item initiators do not correlate well with item responders, and they therefore are not a good predictive measure of dominance in this conference.
- There was a very strong correlation for individuals between their number of responses and the number of items they responded to, which makes it possible to identify participants playing different roles (e.g., dominating one or two items, responding a few times to many, or responding many times to many).
- The participation was surprisingly broad in terms of number of participants.[15]

Findings Generalizable for Archival Description of Computer Conferences

The combination of analytical approaches in this study identified procedures whereby archivists might be able to obtain the following information for any public conference:

1. Key topics in terms of longevity and activity patterns (analogous to looking for "fat files" in archival sampling for appraisal).
2. The degree of internal structure and arrangement of the conference.
3. The breadth of participation in both

[15]Participants join and leave conferences continually. Some also leave the university without resigning from the conferences in which they were listed as participants. As a result, it is impossible to know how many actual participants a conference has at any one time. At the end of October 1991, however, when we began to collect observational data, 360 persons were listed as participants. Of that number, 232 entered responses or initiated items during the observational period. Users of the university's conferencing system also have the options of permanently auditing a conference or temporarily observing one. Neither of these options results in the user's being listed in the participant list, which means that there is also an invisible use and dissemination component to such conferences.

the whole conference and individual items.
4. Dominators or key players among item respondents and initiators.
5. The degree of relationship between item respondents and initiators, which helps archival researchers make inferences about the purposes for which items are being initiated.
6. The difference in usage during different academic periods, which helps archivists make inferences about who the main users might be.
7. To a limited extent, the points at which participants join and drop out of the conference.

Apart from the first two items, none of this information could have been obtained with traditional archival unstructured observation. Nor could the software itself generate this information. Even the first two items could not have been assessed completely by traditional methods, since one must look at the item itself as well as its descriptor to see relators and to identify activity patterns during particular periods.

Conclusions

The study demonstrated how this systematic approach might enhance traditional techniques for the archival description of computer conferences of historical value. This can be done without abandoning the basic paradigms of archival description, which focus on revealing to potential researchers how the material was created and used during its active life, who was important in its creation, and which topics within the collection are particularly noteworthy.

The length and dates of the observation periods worked well as sample sizes, but the data collection process was too lengthy to justify doing manually in a day-to-day archival context. We had hoped to be able to recommend the shortcut of looking only at the item descriptor summaries generated by the system for dominant participants, but we are not able to do so because of the unexpectedly low correlation of initiator data with response data. This was probably the most important contradiction of our research hypotheses. Nevertheless, the data and their analyses yielded much more detailed and precise descriptive information than we could have found using only traditional archival techniques.

To test the generalizability of the methodology, the process described here needs to be replicated on more public conferences of differing natures. Moreover, since the analytical tools used here were exploratory, they collected more data than would be absolutely necessary to an archivist (although they potentially would be very valuable to social science, or historical research). The descriptive methods used in this project should be refined to only the most descriptive, widely comparable, and cost-beneficial techniques, such as the regression analysis of participant and item data and the Friedman AOV between different periods in the academic year. These methods should be used to generate descriptive information that would be included in an archival scope and content note to supplement system-provided item descriptor lists.

We believe that the structured analytical and statistical approach employed in this study has also demonstrated its potential as an enhancement to traditional archival descriptive techniques for other forms of electronic communication. This topic is worthy of further study. We would like to see this approach used as a starting point for the development of a prepackaged statistical or rule-based automated system. Such a system could be used by archivists as a "front-end" in conjunction with the conferencing software, or it could be programmed into conferencing software by conference developers themselves. The system would have potential for being used in both the ap-

praisal and the descriptive processes. It would also have the advantage of being faster, cheaper, and easier to use than the manual process we employed for this study—an essential feature if the system is to be a useful tool for professional archivists who traditionally have little computer and statistical expertise.