Intellectual Access to Archives:

I. Provenance and Content Indexing Methods of Subject Retrieval

RICHARD H. LYTLE

Introduction

AVAILABILITY OF COMPUTERS ENCOURAGES visions of sophisticated information retrieval systems for archives. The visions include national on-line subject retrieval systems using controlled vocabularies and the latest searching techniques. But the visions are shattered by the paucity and poor quality of subject access information about archives: there is little to computerize, and what exists is unimpressive when subjected to the rigors of computer manipulation. Moreover, compiling subject access information about archives can be a very expensive process with uncertain benefits. The challenge, then, is to design for archives information retrieval systems which are both powerful in subject retrieval and cost effective.

System design should be founded upon careful analysis of the characteristics of subject retrieval in archives and upon experimental testing of retrieval systems. The major purpose of this article is to contribute to subject retrieval system design by describing two methods of gaining subject access to archives. These methods are theoretically distinct, but in practice can occur as complementary means of gaining access to archives. The first method, the Provenance or P Method, is the traditional method of archival retrieval, based on principles of archives administration and reference practices of archivists. Subject retrieval in the P Method proceeds by linking subject queries with provenance information contained in administrative histories or biographies, thereby producing leads to files which are searched by using their internal structures. Information in the pure or theoretically defined P Method derives only from what is known *about* the file—the activities of the creating person or organization and the structure or organizing principles of the file itself. The second method, the Content Indexing or CI Method, derives from librarianship but has been applied extensively to manuscript collections, and, to a limited extent, to archives. Subject retrieval in the CI Method matches subject queries with terms from an index or catalog. In the pure CI Method, information is gleaned by an indexer who examines the records; as the CI Method usually is practiced in manuscript collections and archives, provenance-related information is not considered in index-

ing. *Although P and CI Methods do exist as pure types in practice, often they occur as complementary approaches within a given repository. The theoretical distinction is important for analysis, especially for reduction of method characteristics to a form suitable for quantitative study.*

The Provenance and Content Indexing Methods are described below. A companion article forthcoming in the *American Archivist*[1] reports an experimental evaluation of the two methods.

Definition of the Archives System

The following discussion is structured according to the principles of systems analysis, including definition of the system and its components or subsystems, requirements on the system, and responses of the system. The object of the analysis is the totality of archives and manuscript collections in the United States; precisely which materials are included in the system is part of the problem, not a given. Data, admittedly inadequate, are derived from the archives administration literature, the writer's experience, and from the experiment and data analysis reported later. The purpose of the analysis is more to take first procedural steps than to present results.

The archives system is made up of (1) the body of materials contained in the system, (2) the users who make demands on the materials, (3) the finding aids used to access the materials, and (4) those responsible for servicing the materials. As the following discussion indicates, defining the components of the archives system is no simple task.

The body of materials included in the system may be defined in part by the characteristics of the materials themselves. The distinction between archives and manuscript collections is becoming less important with the trend to apply archival techniques of control well beyond the archives of government or other institutions; certainly all materials subject to control by archival techniques should be included in the archives system, without reference to traditional definitions. Moreover, the archives system should include active records of government and other institutions as well as non-current records selected for archival preservation; the traditional dichotomy of retrieval provisions for current and archival records makes little sense in the modern era of records and information management.

The materials in the archives system may be distinguished from materials in other systems. Books, for example, are included in the bibliographic system and are properly distinguished from archives by their focused subject content and their multiple copy distribution. Some materials, trade literature collections, for example, and perhaps some manuscript collections, are not so easily classified. These materials deserve special attention in information system design.[2]

Definition of materials in the archives system may be achieved in part by identifying the system's body of

[1] This article, and the subsequent one reporting experimental results, are based on my doctoral dissertation. See Richard H. Lytle, "Subject Retrieval in Archives: A Comparison of the Provenance and Content Indexing Methods" (Ph.D. diss., University of Maryland, 1979). My advisor, Dagobert Soergel, was a major influence on this work. Also, my intellectual debt to Richard Berner extends well beyond the cited references.

[2] For details on trade literature, see Irene Travis, "Trade Literature at the National Museum of History and Technology," *Special Libraries* 70 (1979): 272–80. Trade literature is an excellent example of material which presents real problems for system definition.

users. Although there are specialized users who present demands for certain kinds of materials, there are many users who ask questions which cross traditionally defined boundaries. A common body of users constitutes the primary reason for regarding archives and manuscript collections as one system. Moreover, active records should be included in the archives system because many users make demands which cross this traditional boundary as well.

Two final components of the system are the finding aids used to access materials in the archives system, and persons who provide reference assistance. The term *finding aid* ordinarily is used only in archives (traditionally defined), but for present purposes it is used for all access devices in the system, including card indexes for manuscript collections as well as administrative histories and inventories for archives. Reference staff may include archivists and other information service personnel.

Definition of the archives system, then, must be accomplished as a first step for information storage and retrieval systems planning; failure to do so will undermine thinking about information systems for archives and manuscript collections. System definition is especially critical for effective planning of a national information system.

Requirements on the System

1. *Who are the Users?* Most archives require that clients register and give some information about their research; usually a record of materials used is kept, as well. These records are kept mostly for security reasons, and archivists have seldom analyzed the information to describe users and their demands on the archives system.

Some data have been gathered for analysis, and a few institutions report regularly on their users. The National Archives has made at least one use study,[3] and the University of Illinois Archives, for example, gives user statistics in its annual report. But there are no synthetic studies, and virtually nothing on the topic appears in the archives administration literature. Gathering and synthesizing existing data is itself a major research task, and it is likely that existing data would be inadequate for cross-institutional studies. What follows, then, is a very impressionistic commentary based on experience, conversations, and years devoted to reading colleagues' annual reports.

Genealogists probably have numerical superiority and item-use superiority, although they may not consume a corresponding proportion of reference staff time. Service to record creators—usually administrators—probably comes next, although this category is appropriate mostly to archives traditionally defined. Scholars probably come next, and within that group it is presumed that historians predominate. Biographers rank high. A last category of user is the casual user, which might include everyone from a chance walk-in to college undergraduates making a pass through the archives on an assignment. In recent years, efforts have been made to widen the user group to include, for example, high school students.

[3] National Archives and Records Service, Office of the Executive Director, Planning and Analysis Division, "A Study of Users of the Records of the National Archives," unpublished, 1976. Available from the Planning and Analysis Division.

The distinction between administrative and scholarly users is pertinent to the traditionally defined archives part of the system. Preserving the record of government activities is the avowed reason for preservation of archives of the state, and the creator, the government functionary, receives high priority in the archives system. It may be presumed that strong creator-orientation was a major factor in development of archival doctrine. As principles of archives administration have been applied to a wide range of non-governmental materials, the primacy of creator-oriented access has been perpetuated.

2. *What Questions Do the Users Ask?* Given the difficulty of defining users of archives and manuscript collections, a description of what questions they ask must be very tentative, indeed. The Finding Aids Committee of the Society of American Archivists has completed two studies on the topic. The first, in 1976, queried archivists about *their opinions* of users' access requirements.[4] The second study, completed but not fully analyzed at this writing, collected data directly from users.[5] Because of a low number of responses and problems with repository selection and response, the results of this study cannot be generalized, but results tentatively indicate the following about users of archives and manuscript collections: (1) most users locate sources through colleagues; (2) users also locate sources approximately equally *via* archivists' suggestions, teachers' suggestions, citations in the literature, and repository guides, but all of these rank well below colleague's suggestions; (3) the *National Union Catalog of Manuscript Collections* is not extensively used; (4) the "most helpful" terms used when searching a repository's finding aids are, first and overwhelmingly, proper names, and second, topics. Despite these studies, however, no significant data exist in analyzable form documenting users' demands on the archives system.[6]

What can, then, be said about users' demands on the archives system? The following continues the impressionistic commentary begun above. Requests received for known documents are rather rare. Usually, these requests derive from bibliographic citations by previous users; and, if the citation is

[4] Kim Efird, "Report on the User Analysis Survey of the Committee on Finding Aids of the Society of American Archivists" (1977). Available from Finding Aids Committee of the Society of American Archivists. This study is mentioned because data were not collected directly from users.

[5] Finding Aids Committee of the Society of American Archivists (untitled: raw data from data on a user analysis survey done by a subcommittee, 1979). Available from the Society of American Archivists. Only 108 useable questionnaires were received from eighteen institutions, and a large percentage of these came from four institutions.

[6] Since appraisal of archives is the process of determining what to preserve for posterity, the appraisal literature might be a source of information about the users of archives and their demands on the system. But the appraisal literature, to quote a recent manual published by the Society of American Archivists, is "disappointing, considering its major significance to archival practice." The manual itself mostly describes appraisal from the viewpoint of the creating organization and characteristics of the records. Due respect is paid to the principle that frequency and quality of use are important to appraisal, and specifically that use statistics should be employed. But beyond these general admonitions the author could not go; the present state of knowledge about use of archives will support no more. See Maynard J. Brichford, *Archives and Manuscripts: Appraisal and Accessioning* (Chicago: Society of American Archivists, 1977), especially p. 2. It is not desirable to attempt to base appraisal entirely on past usage; on the other hand, archivists should study past uses of archives as one criterion for appraisal.

correct and if physical location designators are still valid, response is routine.

Requests specifying names of persons and organizations are said to constitute the majority of requests, and probably this is so even if genealogists are not counted. The strength or degree of name specification varies and the relationship of names in the query to provenance-related information may be an important factor in retrieval.

Subject requests are presented frequently to the archives system. A subject request may be broad or narrow; the breadth of a subject question will be perceived differently in different parts of the system: what appears a broad subject question to a specialized repository might well be a narrow question from the viewpoint of a national system.

Users also may specify date, or may qualify proper name or subject requests by date; most archival bodies and many manuscript collections are easily described by creation dates, and therefore date may be a very important aid to searching. Requests by geographical area are common in some depositories, but probably are less common across all institutions. Sometimes requests by form are used as a means of limiting the search, especially if the form of interest is exceptional in the collection or the repository at hand; a user might surmise, for example, that the information he wants will be found only in diaries.

Requests for records by proper name, geographical area, date, or form may conceal a subject request. Does the user really prefer to ask for documents by name, or does he use proper names only because he has learned that archives access techniques are more effective at retrieval by name than by subject? Berner[7] says that researchers usually request access by proper name because they have associated proper names with activity, and he further argues that proper name is an effective means of subject access. But these assertions have not been subjected to empirical test in an experimental environment.

Browsing is an important characteristic of user behavior. The browser holds an uncertain image of his subject interest; he probably changes his notion of what he wants as his search proceeds. Browsing behavior may be characteristic of historical research, and indeed it may be characteristic of initial phases of all research. Since the requirement to scan files may distinguish one archival retrieval system from another, browsing must be considered in system evaluation. Two cases of users are identified: the user who knows what he wants, does not benefit from scanning files and does so only when required by the system; and the browser, who is uncertain of what he wants and may benefit from scanning files. Apparently browsing and broad subject questions are related, although there is no reason why a user cannot present a broad, precise subject request. Browsing should be distinguished from the use of leads in records, where the user discovers the location of other areas of the file to be searched; for example, in searching a subject file, he finds a reference to related subject heading.

Users generally have modest recall and precision expectations from the

[7] Richard C. Berner and Gary M. Bettis, "Description of Manuscript Collections: A Single National System," *College and Research Libraries* 30 (1969): 410.

archives system. The check on recall—a measure which estimates the degree to which the system retrieves relevant documents—is very erratic. A historian may discover that a document he had not seen invalidates his conclusions in a recent monograph. An administrator may discover that he has repeated a mistake made five years before, because the system failed to retrieve pertinent files. Often, recall failures are less obvious—a scholar or administrator has a "lead" which indicates that he has not seen everything. Probably most recall failures go undetected in the archives system. How much the failures cost is not known.

Precision is a measure of the selectivity of a system. Most users of archives are tolerant of low precision; they are accustomed to poring over large quantities of irrelevant documents. Their patience may derive mostly from acceptance of system limitations, but probably it is related also to a high rate of browsing. Low recall and precision in the archives system in part may be ascribed to low use-rates of archives and low funding levels of archival institutions, resulting in inadequate finding aids. Little effort is devoted to preparation of files, with the result that considerable effort is required when the files are searched.

Clearly, data are needed concerning users' demands on the archives system. Studies by the Society of American Archivists Finding Aids Committee should be expanded. Subject questions should be studied in the context of the educational level of the user, his function (scholar, administrator, etc.), subject question generality, and the browsing phenomenon. Sorting these factors is a prerequisite to evaluation of system response to user demands.

3. *Problems in Defining Users and User Demands on the Archives System.* Although archivists have not attempted to define the users of archives empirically, they know that users change over time. Shifting research traditions bring users with new interests. One illustrative example is the growing interest in women's history: since archivists had no way to predict the widespread study of women's history, they could not incorporate this access point into their finding aids.

Another obstacle to definition of user requirements is users' internalization of system limitations: they request only what they have learned from experience that they can get. A study of user demands on the archives system might reveal overwhelming use of proper name requests, but users might be better served by subject access and might even prefer that approach if given the option. Distinctions should be made between what the user *wants,* what *demands* he places on the system, and what information he actually *needs.* Wants can be determined by canvassing users; demands can be determined by recording users' search requests on systems. But determining what the user needs is a difficult undertaking, requiring in-depth studies of users and relating their information-seeking behavior to task or outcome.[8]

The environment in which user needs studies were developed, information services in mission-oriented research and development organizations, differs radically from the environment of the archives system.

[8] Maurice B. Line, "Draft Definitions: Information and Library Needs, Wants, Demands, and Uses," ASLIB *Proceedings* 26 (1974): 87.

Even within a given area of scholarship, research needs are difficult to assess; the needs are diffuse and the users are unaccustomed to articulating their information needs. Many of these objections apply to general libraries also, but the difficulties may be even more severe in archives. Perhaps the techniques of user need studies are not applicable, or perhaps they are applicable only to a few archival subsystems.

Another obstacle to definition of users and their requirements on the system is resistance by many archivists to social and behavioral science techniques, especially those applied in library and information services. Often this expresses itself in opinions that something "cannot be quantified." A more sophisticated version of resistance is to point out that there are too many variables, known and unknown, to permit practical outcomes from such studies. While there are serious problems in identification of the archives system's users and their needs, empirical studies should be pursued as long as they prove fruitful. Some adjustments of attitudes within the archival profession may be required.

Response of the System

1. *General.* The response of the archives-manuscript collection system to user requirements varies greatly across subsystems. The response in public archives is seen in the development of principles of archives administration culminating in nineteenth-century European theory and twentieth-century American practice. The response in the manuscript collection subsystem, at least until very recently, has been highly individualistic but essentially allied with librarianship. The following discussion describes these subsystem responses. Most of the discussion focuses on the P Method because its retrieval characteristics have seldom been studied. The examination of archival theory is very selective, making no attempt at an historical review.

2. *The Provenance Method of Subject Retrieval of Archives.* Archives administration doctrine emphasizes that records of organizations are created in the course of human activity, and that the resulting organic character of archives is the key to their physical and intellectual control. Provenance is important for control of archives for two reasons: it can be used as a principle of series arrangement which reflects the structure and activities of the creating agencies; and, since agency functions determine the subjects of the records, provenance-related information about archives can be used to gain intellectual control over bodies of records without examining or indexing them.

The closely related principle of respect for original order is, in a sense, a corollary of provenance: if individual files were rearranged in violation of the original order, the principle of provenance would be violated since the organic quality of the records would be disturbed. Yet, the original order doctrine should be considered separately from provenance. Provenance pertains to the activities and structure of the organization which created the record; it does not pertain to individual file order. The principle of respect for original order does pertain to file order, within series. [9]

[9] Richard C. Berner, "Arrangement and Description: Some Historical Observations," *American Archivist* 41 (1978): 179.

Archival doctrine claims more for the Provenance Method than an effective means of intellectual access. The doctrines of provenance and original order state that *information* will be lost if the organic quality of archives is violated. Items in isolation from an archival body lose part of their meaning; the reason for this is that the *file,* not information in the items alone, is related to activity. Thus, the meaning of a file depends on (1) the information in each document; (2) the information derived from the context of documents within a file (file order related to creating activity); and (3) the context of the files—provenance-related information such as that contained in organization charts and administrative histories. Most of the propositions of archives doctrine seem reasonable enough to experienced archivists. None of the propositions has been tested empirically.

The principles of provenance and respect for original order combine into a creator-oriented and document-oriented doctrine. Archives arranged to reflect organizational structure and interpreted in the context of administrative history presumably respond to the information needs of organization functionaries: the creators, their associates, or their successors. Appraisal of archives and intellectual access to archives are founded on creator-oriented analysis and document-oriented description. Physical rearrangement, sometimes proposed to serve users, generally is rejected; the famous late nineteenth-century Dutch manual, for example, specifically rejects the rearrangement of archives to suit the needs of historians.[10] The Provenance Method permeates archival practice, from creation of finding aids to search room reference service.

T. R. Schellenberg outlines a descriptive program as follows: (1) identify the record unit to be described; and (2) enumerate (a) physical qualities, to include cubic feet, medium, etc; (b) substantive qualities, to include the government or other body that produced the record, the functions that resulted in their production, and their subject content.[11] The primary finding aid to archives is the inventory, a shelf list of archives of a specified administrative unit or combination of units. Description relates the documents to the activity that created them, or the purpose they serve. For example, to describe dossiers the archivist explains the action which created the dossiers—to what common matter they relate, or to what function of the administrative body they pertain.[12] Most inventories in large archives specify only at a general level, usually the series.

Subject access in the Provenance Method depends primarily on making the connection between a subject request and provenance-related information. If this step can be taken, the archivist is on home ground; he has a set of functional statements which relate to the structure of the organization and hence at least partially to the structure of the archives. This process has taken him to the record group and ultimately to the correct file which he can use, with assistance of folder lists and the like, in the search for relevant

[10] Samuel Muller, J. A. Feith, and R. Fruin, *Manual for the Arrangement and Description of Archives,* translated from the Dutch by Arthur H. Leavitt (New York: H. W. Wilson Co, 1940, reprinted 1968).

[11] Theodore R. Schellenberg, *The Management of Archives* (New York: Columbia University Press, 1965), p. 101.

[12] Muller, Reith, and Fruin, *Arrangement and Description of Archives,* p. 121.

documents. Most users cannot fully implement this search process themselves, although they may try if the archives has an institution-level guide. The archivist as an intermediary is an indispensable part of the search process in the Provenance Method.

The user is expected to conform to eccentricities of the P Method. The extent of required conformity can be seen in Philip Brooks's excellent book, *Research in Archives, the Use of Unpublished Primary Sources.*[13] First, the potential user is instructed to acquire background in his subject by reading secondary sources. He should be especially careful to acquire knowledge of persons, organizations, places, and events. Researchers must accept the fact that finding aids will not provide access by the desired subject. Indeed, there is no means of accessing the entire archives system; the researcher uses his ingenuity to locate the repository, and then uses guides, other finding aids, and the assistance of archivists to locate collections within repositories. While the P Method does accommodate subject searching by such techniques as subject guides to the National Archives, in general the user must understand the system—almost to the point of becoming an archivist—to retrieve archives effectively.

3. *Enhancement of Provenance Method Subject Retrieval.* The most common method of enhancing P Method retrieval within files is arrangement or rearrangement. Files may be arranged to re-create an original order, to provide order where there was none, or to provide access points different from those of the existing order; in the last two cases rearrangement is similar in purpose to indexing. In the P Method, archives which exhibit an interpretable, usable order must not be rearranged; although this principle is generally accepted, it should be applied with judgment.

Enhancement of P Method power to *select files for searching* has received less attention than have principles of arrangement. In his *Management of Archives,* Schellenberg suggests that P Method subject access can be improved by describing collections with respect to geographic area, broad subject field, and chronological period, using provenance-related information. He recommends analysis with respect to function or activity—for example, agricultural, business, or diplomatic activities. Schellenberg suggests what amounts to a faceted classification, where the facets are human activities. Most important for the present analysis of the P Method, he states that subjects are to be *deduced* from provenance-related information, not from analysis of the files: "Specific subjects, then, should be ascertained deductively...."[14] Schellenberg did not develop these ideas in the literature.

Richard Berner[15] has developed

[13] Philip Brooks, *Research in Archives* (Chicago: University of Chicago Press, 1969).

[14] Schellenberg, *Management of Archives,* pp. 139–40. On the same topic, but largely repetitive of sections in his books, is Theodeore R. Schellenberg, "A Nationwide System of Controlling Historical Manuscripts in the United States," *American Archivist* 28(1965): 409–12.

[15] Richard C. Berner, "Archivists, Librarians and the NUCMC," *American Archivist* 27(1964): 401–9; "Manuscripts and Archives, A Unitary Approach," *Library Resources and Technical Services* 9 (1965): 213–20; "Observations on Archivists, Librarians, and the NUCMC," *College and Research Libraries* 29 (1968): 276–80; "Manuscript Catalogs and Other Finding Aids: What Are Their Relationships?" *American Archivist* 34(1971): 367–72; "Arrangement and Description of Manuscripts," *Drexel Library Quarterly* 11 (1975): 34–55; Richard C. Berner and Gary M. Bettis, "Description of Manuscript Collections: A Single National System."

Schellenberg's ideas concerning P Method subject retrieval. Berner advocates creative use of the subgroup to gain control over collections. For example, when a congressman's papers are received, Berner segregates them according to separate activities of the congressman's life, thereby gaining some immediate control. Berner's ideas parallel similar thoughts in the earlier development of provenance as a method of arrangement of archives above the series level, and he is especially influenced by Oliver W. Holmes's "Archival Arrangement—Five Different Operations at Five Different Levels."[16] Berner further suggests a system in which general subject control is established *via* Schellenberg's notion of activities.

One logical improvement in the P Method is subject indexing of provenance information; the searcher would be aided in making the connection between subject query and provenance information, and then could identify files for searching. To a limited degree this is achieved by indexing descriptions in guides such as the *Guide to the National Archives of the United States.* Schellenberg's and Berner's thinking follow this direction, and Burke and others have advocated indexing finding aids that reflect[17] arrangement by provenance.

None of the ideas for improvement of the Provenance Method's subject retrieval power have been pursued systematically. The subject retrieval process of the P Method must be studied empirically and understood; then it can be systematized and even computerized.

4. *The Content Indexing Method of Subject Retrieval of Archives.* Content Indexing is the practice of accessing collections by *examining the content* of the files. The traditional manuscript collection catalog is the classical application of the CI Method in the archives system. Processors go through collections, piece-by-piece and often without regard to the logic of the collection, applying subject terms which appear as subject added entries in a card catalog. The application of the CI Method in the archives system thus is item-oriented; or, at least, it is not consciously an application of some collective level of description.

The CI Method employs many methods of indexing and many types of catalogs or indexes. The manuscript collection card catalog is only one of these forms. CI Method indexing may resemble back-of-the-book indexing rather than a card catalog, or it may exist as an on-line index file; in either case, the index term will refer to an individual document or group of documents. Moreover, index terms used in the CI Method may be left entirely to the indexer's discretion or they may be suggested by the repository in some form of controlled vocabulary or thesaurus. In the archives system, use of a controlled vocabulary has been uncommon, except in a few repositories following the Library of Congress Subject Headings or the subject headings

[16] Oliver W. Holmes, Jr., "Archival Arrangement—Five Different Operations at Five Different Levels," *American Archivist 27(1964): 21–41.*

[17] Frank G. Burke, "The Application of Automated Techniques in the Management and Control of Source Materials," *American Archivist* 30 (1967): 255–78; "Automation in Bibliographical Control of Archives and Manuscript Collections," in Dagmar H. Perman, ed. *Bibliography and the Historian: the Conference at Belmont* (New York: Clio Press, 1968); "The Impact of the Specialist on Archives," *College and Research Libraries* 33 (1972): 312–17.

of the National Union Catalog of Manuscript Collections.

Content Indexing in principle is very flexible. It can include user-determined access points derived from past user requests, and document-determined access points selected by the indexer. In the archives system, the Content Indexing Method certainly should consider provenance-related information. But historically it has emphasized a rather narrow item-focus, to the virtual exclusion of provenance-related information.[18]

Summary: Two Dimensions of Access to the Archives System.

The preceding discussion has aimed at a *theoretical description* of P and CI Methods. The methods may be summarized as two dimensions of subject access to archives: increasing use of provenance-related context, varying from no dependence (CI Method) to entire dependence on provenance-related information (P Method); and increasing size of the body of materials to be accessed.

A single letter cataloged on its own merits or an item reference in an index employs virtually no provenance-related context and represents the minimum physical size. Folder-level indexing (or some other fairly small aggregation of items dictated by the file) is the next level of physical size which comes to mind; an example (used here because of the experiment reported later) is the Baltimore Region Institutional Studies Center (BRISC). BRISC folder level indexing is an example of larger physical unit but virtually no use of provenance-related information; while BRISC indexes records for relevance to urban studies, context in its P Method meaning is not taken into consideration.

A manuscript collection (or record unit) is a larger physical unit but its description may be based on fundamentally different intellectual foundations. On the one hand, a collection may be described on the basis of notes or index entries made during examination of the contents of the collection, with little or no regard to provenance-related information; in this case, collection level description amounts to a distillation of indexing data derived from lower physical level analysis. On the other hand, a collection level description could be derived entirely from, say, biographical data and knowledge of the organization of the collection involving little or no examination of contents of the collection. A record unit described in the National Archives guide, based on provenance information and knowledge of the structure of the file,[19] is an example of P Method access to a large physical unit.

In any many cases, P and CI Methods are mixed or used as complemen-

[18] The recent manual by Kenneth Duckett is the latest representation of the CI Method approach as described here. See Kenneth W. Duckett, *Modern Manuscripts, a Practical Manual for their Management, Care and Use* (Nashville: American Association for State and Local History, 1975).

[19] Where does proper name indexing fit into this conceptual structure? Personal and organization names are an important part of provenance-related information: *who* participated in the activity; *what* was the result of the activity? Proper names often are used to gain access to subjects, or may consist of most of the same terms as a subject request. Name indexing in the archives system is not peculiar to either method, but the means by which the proper name gets into the system may be classified by method. If the name occurs in provenance-related information or in the structure of the file, it is included in the Provenance Method; if it enters the system by examination of the content of the items, it is considered part of the Content Indexing Method.

tary techniques for archival control and retrieval.[20] For example, archivists preparing inventories may look at documents in a file to some considerable extent and note what they see, while many CI Method indexers doubtless do include provenance-related information in their descriptions. In retrieval, archivists or users may well employ both methods—perhaps the P Method to select files for searching and some file-level CI Method index to search within a file. Moreover, there is no reason to believe that archivists must or should choose between the methods; probably many repositories always will use them as complements in searching. But relative emphasis on the methods will have service and budgetary implications for archival institutions, and hence description of the qualities of P and CI Methods has practical as well as theoretical interest.

Conclusion: A Strategy to Examine Provenance and Content Indexing Methods

The primary reason for attempting a precise distinction between Provenance and Content Indexing Methods is to prepare the research design for an experimental study of the two methods. Thus the foregoing discussion may be viewed as the thought process preceding an empirical study of archival subject retrieval systems. The empirical study has a major objective of further refining conceptualization of P and CI Methods, as well as of providing an evaluation of their relative performance. The wider importance of all of these efforts, as mentioned at the outset of this article, is to support design of better archival subject retrieval systems.

Experimental results will be reported in the next issue of the *American Archivist.*

[20] Interactions between methods also exist (and present special problems for experimental design). For example, one might enter a file *via* the CI Method and then proceed to search in that area of the file, in effect using the P Method. Leads from within files were mentioned above, page 68, in discussing browsing. Leads within files may or may not be method-specific. Information within an item giving new background information is not method-specific, while file structure leads such as cross references definitely are P Method related; some cases resist definition.

RICHARD H. LYTLE is the Archivist of the Smithsonian Institution.