

Data Structure for an Archival Image System: A Proposal

Elizabeth Carley Oddy
Information Transfer Specialist
Syracuse University Kellogg Project

THE PROBLEM

Imagine that you are a researcher, preparing a history of the literacy movement in the United States. You have already combed the published literature for useful information. Now it is time to look at documents created by the various people and organizations who have actually done literacy work. You are aware that, in its Arents Library for Special Collections, Syracuse University holds the papers of Laubach Literacy International and its founder, Frank Laubach, as well as the papers of Literacy Volunteers of America. These are unlike normal library materials in that they are unique collections of personal papers and organizational records.

To study these documents, you would have to travel to Syracuse and request permission to work in the special collections reading room, open only from 8:00 a.m. to 5:00 p.m. on weekdays. You would consult with the manuscripts librarian and look through finding aids to identify the collections you wanted to explore. Finding aids describe the person or organization that created the papers, and list groups of documents within the collection (a folder might contain "Correspondence, January-June 1962" from the Literacy Volunteers papers, for instance). You would not find the author/title/subject lists customary in libraries, nor would you find many things listed individually.

In archives, papers are stored in "archival boxes." A box holds either six or eighteen inches of stacked papers, depending on its type. You would request boxes of papers, then wait for someone to bring them out of the stacks. If the papers had been stored off the premises, you might have to wait a day or two to get access to them. The number of materials you could study at one time would be restricted, and nothing would be allowed out of the reading room. You might be able to bring a quiet portable computer into the room with you; otherwise you would take pencil notes. You could request copies of specific documents, but again you would have to wait for them to be copied by a staff member, and pay a relatively high price.

This state of affairs has existed for a long time. Those who frequently use archive and manuscript collections accept the process and know how to make best use of time on such study visits. But one man dreamed of changing the scenario. Roger Hiemstra hoped in particular to make Syracuse University's valuable collection of adult education manuscripts more widely available and easier to use. The Kellogg Foundation funded his proposal and in September 1986, the Syracuse University Kellogg Project began.

The Model We Developed

What if you could study unique manuscripts while sitting at a computer screen? We, the staff of the Kellogg Project, imagined a system in which you would be able to explore the subject area as well

as the collections of papers themselves, using automated searching techniques. You then could use what you learned in that process to lead you to the groups of documents you sought. You would be able to call up facsimile images of documents, looking like digitized photocopies, to study. If you wanted a paper copy, you could send a command and a good laser copy would be printed for you. When you saw something of interest, you could take notes right away, in another window on your screen, without leaving the research work you were doing. And, best of all, you could do all of this at your own desk.

Although we were unable to make the Kellogg Library and Archive Retrieval System (KLARS) generally accessible over telecommunication networks, our prototype system for computer based archival research incorporated all the other elements mentioned above. In this paper I want to discuss our solution to the problem of providing adequate intellectual access to the document images.

The Key -- Intellectual Access

The access problem turned out to be very complex. We planned to incorporate what had been learned through many years of archival work, but felt we needed more specificity than the traditional emphasis on groups of documents could provide. Systems designed for bibliographic retrieval were specific enough, but weren't a good match with the kinds of documents we were describing. And imaging systems were designed for management of current office records, not research. Somehow we had to bring these three approaches together. Before I describe our solution, let me explain some of the factors that affected our choices.

Differences between Archives and Libraries

Although both libraries and archives collect and store documents for later use, there are important differences in how and why they do it (Rohfeld, Keenan & Oddy). Libraries generally collect published books and journals. Many copies of each book and journal issue are printed, and a variety of libraries may have copies of the "same" ones. Each book or journal is described individually in library records and catalogues, usually by author, title, and subject. The items can stand alone. Cataloguing in one library does not differ in major ways from cataloguing in another.

On the other hand, archives (or manuscript collections--there are technical differences between these terms, but I am using them interchangeably) usually hold unpublished, often unique written or printed material. A particular collection is rarely available in more than one place, and then only through copying each document (on microfilm, for instance). Items are grouped. You would rarely find a single letter identified separately; more likely, correspondence would be grouped chronologically. Papers from one organization or person, one "creator," are kept distinct from all other papers. The organizational scheme used in any given collection of papers will depend on the papers themselves. General principles of organization exist, but are heavily weighted in favor of keeping the "original order" of the collection, as long as it is also "useful".

There are other differences. Books and journal articles are written "about" something, to an unseen audience. Bibliographic tools connect that audience to appropriate materials, using standardized reference points (author, title, subject). Context is presented within documents, along with the ideas

themselves. Archival documents, by contrast, are paper leftovers of business accomplished, life lived. They are reflective "of" activity rather than "about" it, so they resist being categorized into subjects. Researchers are looking in on something that was not intended for their eyes. To understand what they are seeing, researchers must know the context in which the documents were written. Knowledge of context is also required when searching for potentially interesting material. Which person or committee would be likely to have explored the subject you want? When was it being discussed?

Getting to Content

The researcher and the archivist are normally the sources of this contextual information. Researchers are generally held responsible for knowing the subject domain before attempting to work with original materials. Archivists and manuscript librarians have customarily made their considerable understanding of their own collections available to researchers, in interviews as well as in carefully prepared finding aids.

Finding aids list the physical locations of folders and boxes of records, but, more important, they include information describing the history of the papers and the people who created them, the kind of material that will be found in the collection, how the papers are arranged, etc. Researchers and archivists infer the likely location of useful information. Knowing about the organization or person who created the records provides clues to activities, and the documents that result from those activities can be pursued from the clues. The focus is not on the documents but on the activities. (Lytle, Bearman and Lytle, Pugh)

Characteristics of Images

The KLARS design assumed that images of the documents were stored along with machine readable descriptive information. The searching task required the system to locate not just a likely area for exploration, but to produce facsimiles of the documents themselves on the screen. To understand our decisions, it is important to know how images and machine readable text differ.

To create an image record, a page is scanned into the computer in a process very like photocopying, except that the product is an image on a screen rather than a paper copy. Imagine a very fine grid placed over the page, whether it contains text, a drawing, a photograph-- the content doesn't matter. At each square in the grid, the computer analyses whether ink is or is not present. Image scanners vary in complexity, but in the simplest case this is an either/or decision. If there is enough ink in the square, the computer records a full cell or "bit"; if not, an empty cell is recorded. The result is a bitmap of the whole page. The important thing to note is that an image of a page of text is only a pattern of full and empty cells to the computer. The machine cannot "read" the characters, so no processing of the text is possible.

Most people are more accustomed to working with machine readable text. In this case, each character in a text is represented by a standard code, and the codes, grouped into "words", can be sorted, edited, searched etc. The problem in designing KLARS was that, although images of the documents would be available online, the text in those bitmapped images would not be readable by the computer. The original proposal assumed that we would provide electronic access to the

words in the documents themselves, but we quickly realized that if we wanted to be able to search and manipulate that document content, we would have to translate the bitmaps into character data.

Optical Character Recognition

In essence, OCR devices (optical character readers) match patterns found in bitmaps to knowledge of shapes of alpha-numeric characters. Although promising, OCR machines do a far from perfect job of translating and require editing even when given "perfect" copy. To complicate matters, many of our documents were of poor physical quality--fuzzy carbon copies, poor thermofaxes--or were handwritten.

We decided to use OCR, but only on those documents that were likely to come out reasonably free of errors. Approximately 30% of the documents wouldn't be of good enough quality to OCR in the first place. Collection finding aids, generally of good quality and very important to searching, would be edited carefully. Ordinary documents would be left as the OCR translated them. The rationale behind this decision was that system users could always look at the images to read the documents. Machine readable text could then be used primarily as an index to those images. Errors in OCR translation would make this automatic indexing less effective, but would not have the more serious effect of changing the meaning of texts. We wanted to see if the imperfect translations were good enough to be helpful to system users, or whether they were more irritating and misleading than beneficial.

Importance of Context

Part of the logic behind grouping documents rather than treating them individually is that by so doing, some of the context from which they came is kept with items. As I explained above, archival documents do not stand alone like books and journals. It is important to see any given document in light of the activity that produced it. The evidence provided by context can be more important than the actual content of the document itself.

Browsing

Because it is rarely possible to go directly to a specific document when working with archival collections, browsing is an integral part of such research. Previous knowledge, finding aids, and hunches lead to boxes and folders. Because people must browse through these to find what they are looking for, they often stumble across unsuspected resources and links to other boxes and folders. Any system for archival material would have to make such browsing feasible.

SOLUTION

Record types

With the aid of a consulting visit from David Bearman, a nationally respected figure in the world of archive and museum information systems, we decided that a person doing research while sitting at a screen would need at least five kinds of information:

1. Information about whole collections. National databases now contain this sort of record, stored in the MARC/AMC (Archives and Manuscript Control) format. In most situations, one record is created per collection, whether the collection consists of one linear foot of papers or one thousand.
 2. Information about small groups of documents. Finding aids identify collection content by box and sometimes by folder. We knew we would need to get at least to the folder level if we were to bring images of individual pages to researchers.
 3. Information about the players in the adult education world. Biographical details can provide clues to related information. A researcher who knew that a particular person was instrumental in her area of interest would want to identify that person's colleagues, see where the person had worked, when the work was being done, etc. Many useful connections can be made in this way, and can lead to searches through other collections.
 4. Information about adult education organizations and professional associations. This kind of data is similar to that for individuals, and can play a similar role. Knowing about allied organizations, who was active at what time, organizational mission and structure, etc. can lead to new research angles.
- . Because adult education is a field with notoriously overlapping vocabulary, we felt that researchers would need information on use of concept words.

These last three record types are unusual, but are really only new ways of presenting and extending information that has been found useful in libraries and archival repositories under other names, such as authority control, provenance, and vocabulary control. Rather than merely checking to see if we have the proper form of a name, for instance, we would create a more extensive record about the person named and make it available for researchers. We made an effort to record sources of information in case anyone wanted to verify or pursue facts presented in the records. We did not attempt to get the one true fact. If more than one date of birth, for instance, came to our attention, we would simply record all variants and where they had come from.

In archives, knowing about the creators and previous owners of collections is crucial to finding information. Describing the documents can be helpful, but since they were created in the course of normal activity it is most important to understand that normal activity. Bearman and others have discussed extended authority and provenance files and their potential for improving intellectual access to archival material (Bearman & Lytle, Bearman and Szary). Our person, organization and subject records were designed to begin work in that direction.

Matrix

Initially I attempted to attach appropriate fields to each of the five types of records identified above. It then occurred to me that by generalizing some fields and by applying only selected ones from the total set to any given record type, we could design a matrix of fields and record types that would

take care of our requirements in relatively efficient fashion. (See figure 1) Our Indexing Committee fine tuned my original draft through many revisions before settling on the form seen here.

Across the top of the matrix are the five record types: collections, sets (the small groups of documents), persons, organizations, and subjects. Down the left side are the 32 possible fields. Only five fields show up in every record type, and record types employ a maximum that varies from eight fields (for subjects) to twenty-five (for collections). A few fields are designated required; many can be used as needed.

Fields and field labels were generalized in the process of creating the matrix. For instance, the "begin date" field shows the start date of collection coverage, the date of the first item in a set of documents, the birth date of a person, or the founding date of an organization. In cases where this variation is not self-explanatory, brief notations about the differences are included in the matrix. By generalizing, we were able to bring an unwieldy number of fields into a usable set, and to avoid having clusters of similar fields. For instance, collection name, keyword, and name of the person or organization being profiled in a contextual record could all be characterized as main entries. The meaning of that field tag depends on the record type it is associated with.

Each record type shows one "view" of the data, using one subset of the fields in the matrix. For each view we worked out detailed rules for choosing field content, and standards for data entry (Data record views and input standards, 1991).

An interesting thing took place as we thought through the fields in the matrix. Fields added for one purpose took on other roles as well. For instance, most of the "associated" fields were originally included for the contextual record types, persons and organizations. In later discussion, we realized that knowing about associated people, organizations, events, projects, and places might be helpful when looking at records describing collections and sets too, so we extended use of "associated" fields across the matrix.

Subject Access

My original idea for subject records involved collecting information on the origin of concepts, time frame of use, overlap with other concepts, association with people who coined or influenced use of terms, and helpful hints on how to search for the concepts in uncontrolled text. I still think that such files, in a discipline with overlapping and fuzzy use of vocabulary, could be very helpful. However, someone has to do the research to develop the records. I am not aware of any such historical study of vocabulary use in adult education, and we did not have the resources to do it ourselves. We were unable to implement this idea fully.

We still felt that some sort of vocabulary help for researchers would be necessary. By the time we came to grips with this problem, students from Syracuse University's School of Information Studies were already drafting a thesaurus for subject access to documents, a controlled vocabulary for the associated subjects field. We ended up modifying my original design for the subject record type, making it possible to draw subject records directly from our thesaurus, displaying those records in the matrix format.

Data Model Matrix

Field Name	Collection	Set	Org.Vita	PersonVita	Subject
Item number	*	*	*	*	*
Record type	Collection	Set	Organization	Person	Subject
RLIN ID	*	(from coll.)			
Main Entry	Coll. name	(from coll.)	Org.Name	Pers.Name	Keyword
Title	*	(from coll.)			
KLARS source			Who created	Who created	Who created
Physical description	Linear ft.	size & type			
Search strategy hints	*		*	*	Broader.narrow
Bio/historical note	*		*	*	
Summary description	Content	Content	Mission, func. struc.		Scope note
Series	*	*			
Folder title		*			
Set title		*			
Begin date	Coll. coverage	Documents	Founding	Birth	
Source begin date			*	*	
End date	Coll. coverage	Documents	Cessation	Death	
Source end date			*	*	
Form/genre	*	*			
Education				*	
Work history				*	
Assoc. subjects	*	*	*	*	Related terms
Assoc. people	*	*	*	*	
Assoc. organizations	*	*	*	*	
Assoc. events/projects	*	*	*	*	
Assoc. places	*	*	*	*	
Publications	Runs in coll.	Within set	Produced by	Written by	
References to	Pubs based on		Pubs describing	Biographies, etc.	
Location	Syracuse U.	Box, folder	*		
Provenance	*				
Restrictions	*				
Preferred citation	How to cite KLARS	*	*	*	*
OCR'd text	*	*			

Part of the reason for planning subject records was that I never thought we would be able to develop a thesaurus. We are very grateful to faculty and students in the School of Information Studies at Syracuse University for tackling this enormous task. (Allen, 1991)

The Kellogg Adult Education Thesaurus (KAET)

The thesaurus was constructed by "harvesting" terms while looking through contents of the collections, and then grouping and relating those terms to each other. Initial work was done in a thesaurus construction course in the spring of 1990, taught by Elizabeth Liddy and Barbara Kwasnik. That early version was refined and extended during the Fall 1990 semester by one of the students, Eileen Allen. She consulted with adult educators to clarify confusing relationships among terms, added many new terms, and generally brought KAET to a more finished state. Since the bulk of the papers in our collections were created in mid-twentieth century, the vocabulary reflects usage during that period.

The thesaurus is used in two ways in our system, as the source of subject records and as our controlled vocabulary for the associated subjects field in set, person, and organization records. Because collection records were drawn directly from RLIN records, and because RLIN requires the use of the Library of Congress's Subject Headings, we felt we had to use LCSH for collections. On the other hand, LCSH is a very general vocabulary, and for a system that would give detailed access to one general subject domain, it seemed to fit poorly. This produces an anomaly in the system, and is not a particularly comfortable solution.

Automatic vs. Applied Indexing

A companion issue concerns automatic as opposed to applied indexing. In a sense the matrix of record types and fields gave us a framework for applied indexing. It was unusual in that the indexing applied only to groups of documents, rather than individual items, and that it was not particularly topic oriented, but otherwise it is a twist on what has been done before. Automatic indexing of the full texts of documents was the method of giving access originally envisioned. In the end, a combination of the two methods was chosen. The applied indexing would allow human intelligence to help with the characterization of groups of texts; automatic indexing would provide access to content of individual documents when OCR was possible. Neither would be perfect, but these two approaches, when combined with information from the contextual record types and with the traditional provenance-based organization of the documents themselves would, we believed, be effective.

RELATIONSHIP TO FACETED CLASSIFICATION

I have been asked how our record structure relates to faceted classification. I cannot pretend to expertise on the subject, but I will attempt to make a few comparisons.

An intuitive understanding of faceted schemes, by which I mean considering various aspects of the item under discussion, and being able to manipulate those aspects, (search for component bits separately, for instance), bears some resemblance to the way KLARS was set up. However, I don't think the surface similarity holds up well to scrutiny.

First of all, although the different fields could be said to identify different facets of interest to searchers, they are not facets of a subject, rather facets of document or contextual descriptions. They are not designed for purposes of vocabulary control, so much as to systematize the collection of data about records stored in the system. Control was exerted over some of the fields--form/genre, associated subjects, and main entry, for instance. But the overall intent was to give structure to data descriptions.

Certainly our purpose was not to order documents physically into subject oriented classes. Archival information is organized by source and, within such groups, organized alphabetically or chronologically rather than by subject. Our fields were an addition to this provenance structure, meant to expand access possibilities. In an electronic environment we had an opportunity to provide multiple access paths to the data, but we based our work on traditional archival provenance arrangement.

Another difference is that some of our field/facets contained different kinds of data depending on the record type being considered. Although researchers could certainly look for various combinations of fields, some of these field/facets did not have clear, constant definitions across the matrix, but gained definition when they were considered in light of a specific record type. A proper faceted scheme provides constant and clear definitions of hierarchically organized subject facet terms.

IMPLEMENTATION

In early 1987 no commercial product existed that combined the capacity to handle images with full text search capability. We hired a team of programmers to create a customized product that would blend image database software (Plexus Computers' version of Informix) with a text retrieval engine (Personal Library Software). This was an expensive and time consuming process and, for a variety of reasons, it was not successful.

The original hardware and software were not supported by the university at the project's close. In order to preserve some of our work, we developed a HyperCard system, named MacKLARS, on a Macintosh computer. All of the machine readable data was transferred into this system. Unfortunately we lost the images. Faculty, staff, and students in the adult education program, and special collections staff in the library want to use the replacement system.

HyperCard includes, in effect, a very high level programming language (HyperTalk) which makes it possible for "normal" programmers to program multimedia applications on a Macintosh, something that formerly required extensive experience and skill. MacKLARS was created by a regular Kellogg Project staff member, Yuming Tung. The whole process took approximately 5 months.

MacKLARS, like its predecessor, lets researchers search for any machine readable word, whether fielded data or OCR'd text. We used the commercial program HyperKRS as a search engine. It is designed to work with HyperCard software. This combination solves the text retrieval problem that caused us to attempt custom software development for the original system. Although HyperCard

is a tool for building multi-media systems, we were not able in the short time available to us to integrate images into MacKLARS. However, the system is capable of holding them, and it is conceivable that image handling could be added at some later date, given interest, funding and appropriate staff. Unfortunately, the lack of images makes it impossible to fully test our record structure, since the whole point was to bring images of documents to dispersed users sitting at computer screens.

EVALUATION

The ideas presented above were not tested by users in a fully functioning system, although we developed a working prototype. Problems with the design interfered with testing. In particular, the search interface worked clumsily. Handling of phrases was especially awkward, and the system was at times unacceptably slow. The slowness could have been solved with some software and hardware changes, but the search problems grew out of the imperfect blend of the Plexus and PLS components. Each company suggested that the problem lay in the other's code, a frustrating and unmanageable problem.

Applied Indexing -- Groups

We need to know how well the fielding of group data works. Have we suggested too many fields? Which ones actually get used for retrieval? Are there frequently desired hooks to the data that we did not provide? Can users actually learn enough from the group descriptions of documents to find individual pages on a screen without having to browse electronically for an unacceptably long time?

Data input for so many fields is less speedy than would be desirable. It is important for anyone who follows up this work to understand which fields matter, what gets used and what might be left out.

Contextual Record Types

Another set of questions concerns the person, organization and subject records. It makes sense that they might be helpful to researchers, but are they actually used? Have we captured the right information for them, too much, too little? Are users able to make the intellectual links between the contextual information and the documents without the help of an archivist? Are these the most useful types, or are there others that should be considered, in addition or instead? Are these records put to uses we did not anticipate, and if so, what uses?

Creating contextual records is not an easy process, unless they can be based on other, "naturally occurring" data. Each person or organization record required a small research project itself, and the task of creating a substantial number of them is enormous. Knowing which fields are most useful, and finding a satisfactory level of detail are important to the creation process. Our person records, in particular, tend to be very detailed. It would be useful to know how much information is enough. Too much may be more confusing than helpful.

Automatic Indexing with Unedited OCR

There are two ways to think about automatic indexing using unedited OCR. Is it helpful to human beings trying to locate content in individual documents? Could it be useful to a system, giving it data to improve relevance feedback? We made no attempt to look into the second of these questions, although I think there is room for productive research there. I had hoped to test the first question. Unfortunately, the search facility never worked well enough to give this aspect a good trial.

The problems of automatic indexing are well known, and in our case results would be far less accurate than normal. Would it just confuse the issue, under those circumstances? Or would there be enough accurate words to make it better than no access? A major variable is the print quality of input documents. My hunch is that the idea would work well for current documents, laser prints and xeroxes of good type fonts from computers, etc. It is less clear whether it will work for the kinds of documents we were capturing, created by typewriters and duplicated by carbon paper, thermofax, and early photocopies. Of course, current documents are often available in machine readable form, making OCR unnecessary. The question of how users would react to such a rough indexing tool is an important one, and not being able to answer it is a great frustration.

System as a Research Tool

Everyone who saw the system was impressed by its potential. The problems were primarily technical ones. Conceptual work was all there. If we are able to complete MacKLARS by adding in the images, we will finally be in a position to see how users react to the ideas. That system, far faster, easier to use, and containing a full user-friendly search mechanism, will be less intrusive. Using KLARS, the ideas got lost behind interface problems.

IMPLICATIONS

Without data from users, discussion of implications is rather more theoretical than I would wish. Here, then, are some tentative conclusions:

As image systems become more widespread, mechanisms for giving access to the images will become refined. The mixture of text retrieval and image management that we implemented is a harbinger of things to come. Commercial systems that combine these elements are just beginning to arrive on the scene. But even when they are widespread, people will still need to decide how to organize their data. It is here that our work may be able to show the way.

David Bearman has argued persuasively that the incorporation of MARC/AMC into the MARC family of formats, and perhaps eventually into a unified MARC format, has the potential to revolutionize the way the information world does business. Perhaps individual cataloguing of documents, divided by physical format, is not in the long run the best way to do business in an exploding information world. And perhaps we are moving toward a world in which there is more than "one right way" to present information; perhaps it is time for multiple authorities, multiple intellectual access points to information. Computerized systems allow many more options; perhaps its time to change our understanding of what authority control means (Bearman 1989a, 1989b).

Our solution to the problem of giving access to specific images while making use of traditional archival processes may provide implementation suggestions to others. Important elements of the solution include:

1. Using a form of applied indexing on groups of documents, and combining that with automatic indexing of individual document content to give both breadth and specificity to online document access.
2. Creating records that describe the context in which documents were created (people, organizations, subjects), and keeping those separate from but available with records that describe the documents themselves.
3. Using unedited OCR'd versions of text from images as rough indexes to those images.

These methods might prove useful in both business and research environments.

SUGGESTIONS

The link between our system and the MARC/AMC format is probably not close enough. Certainly collection records and probably set records as well could be framed in MARC style. David Bearman and others have urged the extension of authority files into reference files that could be made available to scholars as well as processors (Bearman and Szary). As one of the designers of the AMC format, Bearman tried to make the container broad enough to hold such files. Anyone doing additional work along the lines we have begun might want to consider whether it is possible to develop necessary structures for contextual fields within the AMC framework, or the general MARC format, when that is put into place. Staying within MARC boundaries would open up possibilities for sharing data in the long run, particularly important when images as well as ascii data can be shared easily over networks.

Although we were lucky enough not to become a test case, anyone working in this area needs to keep a weather eye on developments in copyright law. The system we dreamed of could have become a case for debate between those supporting ownership of intellectual property rights and those supporting improved access to research materials. It is difficult to find a mutually satisfactory path between these competing interests, and copyright law is only beginning to deal with the new wrinkles brought by the advent of image systems. In the next few years there will no doubt be test cases in which some of the rules of the game are set down. Until then, caution is in order.

CONCLUSIONS

It takes time for the possibilities inherent in new technologies to become evident. Our initial human tendency is to do the old jobs in the new environment for a while; only later do people begin to see options that make formerly unimaginable things possible. I think that image systems will eventually move in fresh directions, and new techniques will be needed. Perhaps what we have done will suggest welcome ideas to those who follow.

REFERENCES

- Allen, Eileen, *Kellogg Adult Education Thesaurus*, 2nd ed. Syracuse University Kellogg Project Technical Report No. 5, April 1991. (Available through ERIC)
- Bearman, David A and Richard H. Lytle, "The power of the principle of provenance", *Archivaria*, v. 21 (Winter 1985-86), p. 14-27.
- Bearman, David and Richard Szary, "Beyond authorized headings: Authorities as reference files in a multi-disciplinary setting", *Authority Control Symposium*, ARLIS Occasional papers #6, ed. Karen Muller (Tucson: Art Libraries Society of North America, 1987), p. 69-78.
- Bearman, David, "Archives and manuscript control with bibliographic utilities: Challenges and opportunities", *American Archivist*, v.52, no.1 (Winter 1989), p. 26-39.
- Bearman, David, "Authority control issues and prospects", *American Archivist*, v.52, no.3 (Summer 1989), p. 286-299.
- "Data record views and input standards," *Kellogg Library and Archive Retrieval System Document Capture Manual*, Syracuse University Kellogg Project, May 1991, Appendix A, p.104-121. (Available through ERIC)
- Lytle, Richard H., "Intellectual access to archives: I. Provenance and content indexing methods of subject retrieval", *American Archivist* v.43 no.1 (Winter 1980) p. 64-75.
- Lytle, Richard H., "Intellectual access to archives: II. Report of an experiment comparing provenance to content indexing methods of subject retrieval," *American Archivist* v. 43, no.2 (Spring 1980) p. 191-207.
- Matters, Marion, *Introduction to the USMARC format for archival and manuscripts control*, Society of American Archivists, 1990.
- Oddy, Elizabeth, *Kellogg Library and Archive Retrieval System: Design and development*, Kellogg Project Technical Report no. 6, Syracuse University Kellogg Project, August 1991. (Available through ERIC)
- Pugh, Mary Jo, "The illusion of omniscience: Subject access and the reference archivist", *American Archivist* v.45 (Winter 1982) p. 33-44.
- Rohlfed, Rae, Terrance Keenan and Elizabeth Oddy, *Introduction to archival research: How are libraries and archives different?*, Syracuse University Kellogg Project, May 1990. (Available through ERIC)

