The current issue and full text archive of this journal is available at
www.emeraldinsight.com/0956-5698.htm

REPUBLISHED SELECTION OF ARTICLES TO COMMEMORATE THE ANNIVERSARY

Records management and data management: closing the gap

John McDonald

Automated Information Systems Division, Government Records Branch, National Archives of Canada, Ottawa, Canada

Abstract

Purpose – The purpose of this paper is to explore the information landscape of organizations by focusing on the evolution of the fields of so-called records management and data management.

Design/methodology/approach – The author draws on his personal experience with the National Archives of Canada.

Findings – Records management and data management quite literally mean the same thing. There is no "gap", as indicated in the title. The only gaps that exist are in the perceptions of what each concept means and the functions and status of the information jurisdictions that have claimed each for their own.

Originality/value – The paper recommends an integration of what has been perceived to be the disparate fields of records management and data management, finding that records or data should be managed from a global and corporately defined perspective

Keywords Records management, Archives, Document management, Digital storage

Paper type Conceptual paper

The other day I was at a meeting with several government officials to discuss the establishment of a retention, conservation and disposition schedule for a large, complex, automated information system. Around the table were users, systems people, the head of data administration and the departmental records manager. The first thing that intrigued me about this meeting was that the records manager and the systems people (including the data administrator) had never met each other, even though all had worked in the department for at least two years. For someone who was concerned with launching an initiative that would depend on the cooperation of these areas I was more than a little shaken.

The conversation that followed was even more intriguing. The records manager claimed that, pursuant to departmental and government-wide policy, he was responsible for all the *records* in the department, including electronic records. Given this understanding he felt that he should serve as the principal contact point, coordinator and overall authority for any project that involved the establishment of schedules for records in automated information systems. The systems people were a

This is the text of a talk first delivered to the Society of American Archivists, Annual Conference, Atlanta, Georgia, October 1988.

This article was originally published in *Records Management Journal*, Vol. 1 No. 1, pp. 4-11, and has been republished as part of the journal's 20th anniversary commemorative issue.

Records Management Journal
Vol. 20 No. 1, 2010
pp. 53-60

0956-5698
DOI 10.1108/09565691011039825

little perplexed because they had assumed that the records manager only looked after paper records. And anyway, the electronic information in computers wasn't a *record* – it was data – so it didn't count. Furthermore they felt that anything concerning the retention and disposition (including archival transfer) of data was up to the users.

Needless to say, the users were becoming somewhat perplexed by the jurisdictional battle that was unfolding before them and, since they had never had to think about the retention and "archiving" of the data for which they were supposedly responsible, concluded that they really didn't want to be at this meeting.

That experience and others like it have taught me the importance of understanding the information management landscape of an organization before undertaking an initiative to develop retention and disposition schedules covering all recorded information, be it in a system or in the organization as a whole.

The purpose of this article is to explore this landscape by focussing on the evolution of the fields of so-called records management and data management.

Records vs data

While emphasizing the gap that has existed between our perceptions of the fields of records management and data management, I would also like to explain how the gap is closing, particularly in light of the evolution of networked office automation systems. Finally, I will explain how the requirement of archives for institutions to establish retention, conservation and disposition (including archival transfer) plans and specifications for recorded information can serve as a catalyst for bringing these two fields together.

In exploring this landscape it might be useful to begin by determining if we really understand what is meant by "records management" and "data management". What is a "record" or "data"? In the Canadian federal government the definition of a record can be either the media upon which information is recorded or information stored on various media[1]. Regardless of the perspective, "record" refers to any recorded information regardless of media. Therefore it follows that records management is the application of the management principles of planning, organizing, directing and controlling to the recorded information holdings of an organization.

What is (or more properly "are") "data"? According to international definition[2], data are the representation of facts, figures and concepts in a manner suitable for processing, interpretation or communication by *human* or *automated* means. In other words data refers to any recorded information regardless of media. It also follows, therefore, that data management is the application of the management principles of planning, organizing, directing and controlling to the recorded information holdings of an organization.

Records management and data management quite literally mean the same thing. There is no "gap" as indicated in the title. The only gaps that exist are in the *perceptions* that we have of what each concept means and the functions and status of the information jurisdictions that have claimed each for their own. Records management, for instance, immediately seems to bring to mind the image of underpaid file clerks looking after paper records in file folders, while data management seems to bring to mind highly paid computer systems specialists or data administrators planning and designing complex computerized databases. The distance that exists between the two in terms of their skills, their tools and

techniques and their place in the organization, have reinforced our perception that records management and data management are two distinctly separate things.

This fragmentation is a reality that continues to frustrate those information practitioners who are searching for a more holistic approach to their job – one that emphasizes the integration of the tools, techniques and skills of professionals representing a range of information disciplines. It is a reality that continues to frustrate those senior executives who are concerned with introducing a corporate approach to the management of their information holdings. And finally it is a reality that continues to frustrate those archivists who are concerned with obtaining a comprehensive view of the information universe of a given organization before installing mechanisms that will provide for the systematic transfer of archival records, regardless of their physical form.

The differences that exist between the two worlds of records management and data management (and here I am speaking of data management as it has emerged from the database administration and data administration functions of the information systems areas of organizations) can be seen in the fragmented legislative and policy framework that governs the management of information in organizations. In the Canadian federal government, for instance, information laws such as the Access to Information Act, the Privacy Act and the National Archives Act, refer to all recorded information. However, the policy framework and the standards and implementation guidelines that give expression to these laws are divided between those that govern hardcopy and those that govern electronic data processing or EDP records.

The Records Management policy[3] for instance, addresses the management of all records except computerized information. Policy direction for the management of computerized information or data is buried in EDP policies that govern the acquisition of computer technology and the development of automated systems[4]. While a draft policy covering the management of all information holdings[5] is being developed it is still in draft form and is not without controversy – and for a good reason. If approved, it could impact substantially on such information jurisdictions as records management and information systems. In assigning responsibility for the policy to a senior official, these jurisdictions would be required to coordinate their efforts to serve a corporate approach to the management of information. Above all it would require institutions to fix authority and responsibility for the life cycle management of all information – a requirement that could be difficult to implement given the lack of clear definitions of the roles and responsibilities of various key players.

The fragmentation that exists at the legislative and policy level is also reflected in the backgrounds and skills of the professionals who support the records management and data management functions of organizations.

Traditionally, records managers have been viewed as administrators of paper records – individuals who have emerged from the mail room, who have limited skills, and who are essentially administrators of hardcopy filing systems. Generally, they are located at a low level within the organization. They are also poorly educated. In the Canadian federal government, for example, a minority of records managers have post secondary education. In fact, any university level courses on records management are often imbedded within library science programs. At the community college level, multi-year records management programs exist but are not widely known (let alone recognized). A particularly disturbing observation is that only a few records managers

have received extensive management training. In federal government they have advanced through long years of experience, hard work and personal initiative. They have done so in an environment where users (too many users) share the view that records managers are elevated clerks who, though undoubtedly skilled in what they do, have little bearing on the day to day operations of the program activities of their organization.

Those responsible for the management of data, however, are generally perceived to be graduates from computer science courses who have mastered the technical skills required to build complex databases and systems. In addition to the availability of graduate and post graduate programs in computer science, these individuals (they include those systems development analysts, data administrators and others who are responsible for planning and otherwise managing the data resources of the organization) can draw from a rich set of related programs that emphasize the application of computers in the design and administration of systems – business administration, engineering, library science, to name but a few. Information systems professionals are often located at middle and even senior levels of the organization. Although some (and I believe their number is decreasing) will often defer issues surrounding the definition and retention of data *entirely* to users (i.e. they only provide technical solutions to user requirements), others have been taught the importance of strategic information planning, the need to manage data resources from a corporate approach, and the need to integrate the tools and techniques of systems development and data management into the organization's management of its business functions. In mature organizations, information systems professionals (particularly those that demonstrate strong information planning and data management skills) are being recognized as an important element in the achievement of corporate goals. Such individuals have gained this recognition through, in part, a broad education background that extends beyond computer science to management and ancillary fields such as business administration.

Tools and techniques

Although the disparity that exists between the education and skills of information systems professionals and records managers is significant, an area of even greater disparity lies in the tools and techniques that each uses to manage the information for which they are responsible.

Records managers, for instance, have generally been concerned with the management of documents – unstructured text created in a non-repetitive manner to serve one or a wider variety of uses. Examples include memoranda, reports, and correspondence. Information systems professionals, however, have generally been concerned with the management of data – structured information created on a repetitive basis and processed through (generally) pre-defined systematic steps to produce a predefined information product such as a license, cheque, etc.

In order to provide a corporate perspective to the management of documents, records managers have developed tools such as the subject file classification system and the retention and disposal schedule.

In managing data in a systems environment, however, information systems professionals have developed tools such as the systems development life cycle, systems development methodologies, tape library management systems, and data dictionaries.

The data dictionary, for instance, is a tool that serves as a central storage of data (descriptions, attributes, and relationships) about data used by computerized and/or manual systems. If used properly, it captures the corporate data model that can be used to build applications that are consistent in design. At a broader level, strategic information planning tools and business systems analysis tools are being used to ensure that a corporate perspective is adopted for all systems and database development activities. The introduction of sophisticated automated systems development tools (e.g. CASE tools) combined with the emergence of a technically literate user population, have served to transform the entire way that systems professionals do their work.

Nevertheless, although the evolution of data management concepts and the emergence of strategic information planning have promoted the adoption of a corporate approach to the management of an organization's data, they have yet to account adequately for those other information holdings that reside in office filing systems or such specialized areas as map libraries, photo libraries or even standard libraries. Many of the strategic information plans that have been prepared in organizations have failed to account for these important components of the information landscape.

How then can the gap between these disparate functions be closed? I would suggest that there is one area in which the gap will be required to close whether the organization is prepared for it or not. And that is in the area of networked office systems, particularly those systems that comprise several or hundreds of microcomputer workstations all linked electronically together to permit users to create, transmit, store and dispose of recorded information some of which may never reach paper form. As large numbers of these networks are installed, organizations have begun to discover the challenges of managing electronic information much of which is in the form of documents. How many of those who use personal computers have ever lost an electronic document – they couldn't find it, they couldn't remember what it was called or what version it was, they had accidentally deleted it, or the power went off when they were going to save it.

If so, then it is not difficult to imagine the large organization which must cope with hundreds of personal directories containing hundreds of documents with 8 character titles. How can some measure of corporate control be applied to ensure the ongoing care of the electronic corporate memory of the organization?

On a slow but increasing scale, both the so-called systems integrators (i.e. the vendors of networked office systems) and the institutions that are acquiring and using these systems are recognizing the need to apply records management concepts and practices to solve the corporate information management problems of what is essentially (at least at the outset) a *document* based environment.

Over time, however, and as these systems mature, the problems associated with managing *data* either in local "homegrown" databases or in applications that involve the downloading of data from corporate databases will give rise to the need to introduce the principles and practices of data management. It is at this stage in the evolution of networked office systems that the *document*-oriented world of records management will meet (and, hopefully, not clash) with the data oriented world of data management.

The corporate approach

If this is to occur in a planned way, however, steps must be taken to understand how the tools and techniques of each world can be linked or even combined to support a corporate approach to information management across the office systems landscape. Relationships between file classification systems and data dictionaries, between retention schedules and update/delete cycles, between document protection and data integrity procedures and so on, must be understood and rationalized. On a broader scale, a policy framework that assigns responsibility and accountability to each of the key players in both worlds must be established. And above all, given the dominant role of the user in the management of information generated in these systems, considerable effort must be made to develop a comprehensive set of tools to permit users to manage what they will perceive to be their information, albeit in a corporate context.

The effective application of document management and data management concepts in this complex environment can occur only if an institution has prepared a comprehensive *plan* – in effect an information plan that addresses the introduction and use of both information technology and information management practices. Such a plan must be developed in accordance with a clear set of *requirements*, requirements that are expressed in a manner that is independent of technologies requirements that serve as a base for the development of functional specifications for the management of various information types associated with particular application environments – both document based and data based.

To those who suggest that this is an ideal that is years or even decades away, I would suggest that the sheer necessity of users (particularly those who function in large work groups) to bring some semblance of order (and preferably corporate order) to their increasingly sophisticated information holdings will oblige both the vendors of office systems technologies and the information practitioners of institutions to respond.

Although the office systems environment should be a catalyst to promoting the greater integration of records management and data management concepts and techniques, I would suggest that an additional and equally important catalyst can be found in the efforts by archivists to establish agreements or schedules that govern the ongoing systematic transfer of archival records to the archives. Schedules describe the length of time that institutions are to retain their information holdings, the manner in which their physical and intellectual integrity is to be conserved during this retention period (particularly important for electronic records) and the manner of their disposition (i.e. destruction or transfer to the archives). A schedule or agreement however, must be based on a comprehensive understanding of the information landscape of a given program activity. This is particularly important for archivists who are concerned with appraising records within the context of both the program activity as well as the interrelationships that exist among records.

In the Canadian federal government, institutions are required to prepare descriptions of their information holdings such that a comprehensive, albeit generic view of the information landscape can be obtained. The identification requirements or descriptive standards that the archives develops to assist institutions in building this comprehensive view can serve as a major catalyst in encouraging departments to develop comprehensive directories to their holdings – high level directories that point to otherwise disparate finding aids such as subject file systems, data dictionaries, and

so on. These directories and related finding aids can serve as a basis for the establishment of retention and disposition schedules.

In building retention and disposition schedules or agreements, organizations must not only identify their holdings, they must also set meaningful retention periods. The establishment of retention periods, however, requires an understanding of the interrelationships that exist among records (regardless of their physical form) and the context within which these records are being used. While the setting of retention periods for data in automated information systems must be based on an analysis of all of the components of a system (inputs, processes, outputs), a schedule is not complete until all of the records that pertain to the management and administration of the system, including those in subject file systems, are included. Similarly, it is not sufficient to set retention periods for the documents (electronic or hardcopy) in a subject file system without also setting retention periods for the data (hardcopy or electronic) in an automated information system.

Although Allan Kowlowitz has demonstrated that such a comprehensive approach to scheduling is difficult to accomplish[6], particularly for those systems that cross program and even organizational boundaries, the need to reach towards a comprehensive approach to the establishment of retention periods remains.

The comprehensive identification of an organization's information holdings and the comprehensive analysis that is required for the establishment of retention periods must be followed by a comprehensive approach to the conservation of information holdings for the duration of the retention periods described in the schedule. This is particularly important in the case of those records that have been assessed as having archival value.

If these records comprise both data and documents (either hardcopy or electronic), it will be incumbent on the institution to adopt an integrated approach to the development of technical and procedural specifications for the conservation of its corporately valuable information holdings, particularly those recorded in electronic form.

Finally, the scheduling function will require organizations to specify the disposition of records, either through destruction or transfer to the Archives. Again, organizations will be required to adopt an integrated approach if they are to ensure that both data and documents are properly destroyed/deleted or transferred, on a scheduled basis, to the Archives.

The identification, retention, conservation and disposition (IRCD) of recorded information, as specified in a schedule, will require a management framework that fixes accountability and responsibility across the organization. As a result, the steps leading to the establishment of retention and disposition schedules or agreements can serve as a catalyst that forces an organization to explore how the two worlds of data management and records management can be combined. On a broader scale and in accomplishing this task, the organization can begin to learn how a more comprehensive approach can be adopted to the overall management of its information. This becomes particularly important when the organization recognizes that the scheduling requirement can be best addressed when new programs are being established and new systems are being developed, preferably in accordance with a strategic information plan.

For its part, the National Archives of Canada is moving away from a piecemeal approach to acquiring archival records (based on schedules submitted in accordance with the agenda of records managers) to a planned or strategic approach in which the establishment of agreements is based on the development of scheduling plans with institutions. On more than one occasion officials in institutions have suggested that by addressing the IRCD issue, their institutions will be well on their way to integrating the information management practices of their organization.

While it is true that the scheduling function may have little impact on the most important elements of information management (i.e. identifying information needs, developing information plans, etc.), the role of the archives as a catalyst in integrating what has been perceived to be the disparate fields of records management and data management, should in itself cause organizations to take a second sober look at what it is they are managing, and why it is so necessary that the "it" (record or data) be managed from a global and corporately defined perspective.

Notes

1. The definition of "record" is based on the *Access to Information Act* and the *National Archives of Canada Act* (note English and French versions).
2. The definition of "data" is based on definitions produced by the International Standards Organization; see *Glossary of Terms*, Government EDP Standards Committee, Ottawa, 1982.
3. *Chapter 460 – Records Management*, Administrative Policy Manual, Treasury Board Canada, Ottawa, 1983.
4. *Chapters 435, 436, 440 (EDP)*, Administrative Policy Manual, Treasury Board of Canada, Ottawa, 1978.
5. *Management of Information Holdings*, draft Treasury Board policy, Ottawa, August 1988.
6. A. Kowlowitz, "Hands up, you're under arrest: appraising criminal history data in the age of the electronic case file", paper presented presented to the Annual Conference of the SAA, 6 September 1987 (A. Kowlowitz, New York State Archives and Records Administration).