

Archival Description of Electronic Records: An Examination of Current Practices

by **JEAN E. DRYDEN***

Résumé

Cet article décrit les résultats d'une enquête menée dans le domaine des pratiques descriptives actuelles des dossiers électroniques en espérant que les pratiques présentes nous renseignent sur le développement des règles de description. Les résultats de cette enquête se sont révélés d'être quelque peu décevants quant à la problématique de la recherche initiale. Toutefois, l'analyse des outils de description actuels éclaire plusieurs questions relatives aux documents électroniques, à la description, et à l'administration des archives; l'article identifie également un certain nombre de secteurs où une recherche plus approfondie est encore requise.

Abstract

This article describes an investigation into current descriptive practices for electronic records, in the expectation that current practices would inform the development of descriptive standards. The rules of the investigation proved to be somewhat disappointing in terms of the original research problem. The analysis of existing descriptive products, however, reveals some interesting insights into a number of issues relating to electronic records, description, and archival administration; and identifies a number of areas where further research is required.

Archival description of electronic records has been but generally addressed in the archival literature, and much of the current discussion has been, of necessity, speculative.¹ Archival description is rather more complex than commonly represented; in order to examine this area in detail, and to move the discussion from the realm of theory to empirically tested and accepted professional practice, several approaches are possible.

As in many other professions, archival practice develops in an evolutionary, not a revolutionary manner. For a number of reasons, progress is slow. Although electronic records issues have been on the agendas of some institutions and individuals for more than twenty years, the development of appropriate practices for the description

of electronic records has barely begun. The profession is exploring two particular initiatives: the development of specific data content standards within the framework of the *Rules for Archival Description (RAD)*, and a widely dispersed (and fragmented) examination of metadata.

The study described in this article, while focusing on the *RAD* project, attempted to combine two approaches to the problem. In developing the most appropriate professional practices to describe electronic records, one obvious area to examine is that of current descriptive practices to see what works and what must be changed. Another, related approach is to develop a method or procedure for dealing with the new problem and to test it.

This study, therefore, had two purposes: first, to see how archivists are currently describing electronic records and to examine the elements of these descriptions for completeness, relevance, and commonality within and between institutions; and second, to find examples of actual electronic records in archival custody that could be used to test the proposed² descriptive rules in the *RAD* chapter on electronic records, and to illustrate the rules once they were approved. The task seemed straightforward; in fact it proved to be surprisingly complex. Moreover, while it did not produce the expected or desired results, the actual results were nonetheless very revealing. While it provided relatively few useful examples of electronic records descriptions, the exercise revealed valuable insights into a number of issues relating not just to electronic records, but also to broader issues of description and archival administration.

There was not much choice in sources of relevant examples and related procedures. Relatively few archival institutions have any electronic records holdings; even fewer have devoted sufficient resources to this area so as to develop any significant body of experience with them. Thus, the sample was necessarily limited to institutions that would have not only descriptions, but also documented procedures for their preparation. The institutions chosen were the National Archives of Canada (NAC), the United States National Archives and Records Administration (NARA), and the New York State Archives and Records Administration (SARA)--three institutions with extensive experience in the management of electronic records, and with very similar mandates: all are the official repositories of the records of large and complex governments, with significant responsibilities for both records management and archival functions.

Initially they were asked to provide examples of descriptive products for as great a variety of electronic records as possible, plus the institutional procedure manuals and related standards for the structure and content of these examples. Of particular interest was the way the technical aspects of the records were described. Because these examples were going to be "translated" into *RAD*-compliant descriptions, it was important to have descriptions at the summary³ in order to examine the ways in which levels of arrangement were represented in the descriptions. Included was a general request for any other procedures, forms, or other documents that provided information used in the descriptive practices and products of the institutions.

The responses varied with each institution, depending on the availability of institutional procedures, and on how broadly each respondent defined description. The range was indeed varied, and included appraisal reports and guidelines, documentation accompanying the records in their transfer to the archives, general

information on the electronic records programme, records inventory forms, and published guides.

Upon examining these documents, it was clear that this seemingly straightforward task was not going to be straightforward at all. The sample of “complete” descriptions (that is, descriptions of records that were considered to be fully processed) was very limited; much descriptive work was still in progress or awaiting attention as part of the backlog. Furthermore, the range of electronic records systems was not represented in the examples. The only “complete” descriptions were of so-called “flat” data files; there were no complete descriptions of more complex electronic records from office systems (word-processing, spreadsheets, electronic mail, etc.), variously structured databases (i.e., hierarchical, network, or relational), or more complex systems such as geographic information systems.

Clearly it was not going to be possible to investigate current descriptive practices simply by examining completed descriptions. It thus became necessary to try to construct descriptions of “real” records from the information captured in the earlier stages of archival processing. Many of the related documents and forms already sent were very useful in this process, providing a rich store of descriptive information. The array differed for each institution, however, because each had interpreted description differently, and had provided procedures and related information for different stages of archival processing. In seeking out common descriptive elements, it was important to have as much similarity as possible in the range of documents from each institution. In addition to forms it was important to understand each institution’s processes and procedures. Thus, it was necessary to go back to each institution and request that it fill in the gaps if possible.

After this second stage of data gathering (followed up where necessary with interviews), there were samples of seven categories of documents (a sample could include any combination of forms, procedures, and/or examples):

- Records inventory or schedule,
- Appraisal report,
- Accessioning procedure,
- Descriptive rules/standards,
- Descriptive record,
- Finding aid, and
- Published guide.

The chart on page 108 shows the categories and the particular documents each institution uses/produces in each category.

Considerable time was spent reading, sorting, and analyzing the array of information collected. It was clear that there was no comprehensive body of elegant descriptions and accompanying procedure manuals documenting current practices. However, the data collected from forms, procedures, examples, and conversations provided a considerable amount of information about how much descriptive activity is taking place, the records being described, the nature of descriptive products being produced,

the descriptive standards being used, the state of descriptive procedures, and the administrative context in which these activities take place.

The respondents in these institutions all commented (somewhat apologetically) that very little description was going on. That is true if one defines description as the function that takes place after records have been accessioned and arranged. If, however, one takes a wider view of description and considers all the documents in the chart as descriptive products, it is clear that a great deal of descriptive activity is happening, albeit in a fragmented way. Two broad categories of descriptive products are represented here: those that are traditionally considered to be the products of description (items 4-7), and those that are the products of other functions but contain significant amounts of descriptive information. In this latter category, obviously accession records were particularly useful, since the accessioning process requires a minimal description in order to establish basic archival control over the new holdings. Not surprisingly, appraisal reports were also a particularly rich source of descriptive information. Records inventories and schedules prepared before the archival processes (traditionally defined) begin provide much valuable information. Taken as a whole, however, this range of products is not integrated across functions or within the institutions.

As the first attempt at data gathering had revealed, the range of types of electronic records being described is very narrow. There appear to be two main reasons for this limited range: relatively few such records are yet in the custody of the archives, and those that are have not been described in any detail.

The descriptive standards used are either data content standards or data structure standards such as MARC(AMC) that are being forced to serve as data content standards.⁴ They represent three types: those developed in-house, those adopted (to a greater or lesser extent) by the profession, and in-house adaptations of recognized standards. Where data content standards serve as the basis for descriptive rules, the standards vary depending on the level of description. The standards are generally derived from two sources: *Archives, Personal Papers and Manuscripts* for series level descriptions, and Chapter 9 of *Anglo-American Cataloguing Rules* (Machine-readable data files) for descriptions of media-specific material at a lower level, often as a finding aid. There is a well-developed body of practice using the rules for machine-readable data files, drawn from the experience of data libraries. While these rules work well for data files (so-called flat files), they are not suitable for more complex records. In other cases, where rules are being used at a higher level to describe multimedia fonds or record groups, those for textual records have been adjusted to accommodate electronic records (e.g., NARA's 1993 changes to their Format X rules developed in 1991).

The study found that procedures for archival description are not comprehensive. For some functions there are numerous examples, but no guidelines or procedures standardizing the practice. For others, the procedures have an interim status, awaiting the testing of a particular procedure or the development and implementation of new automated systems. For example, the accessioning procedures for two of the three institutions are designated "interim."

The administrative context in which this work takes place differs along national lines. All three institutions are large complex organizations in which it is difficult to

maintain consistent and complete communication between and among all the parts. Both American institutions have established separate Centers for Electronic Records, in order to concentrate resources and expertise on this medium. The National Archives of Canada was similarly organized until 1986, when the institution underwent a reorganization along provenance/functional lines and responsibility for electronic records was integrated into the areas responsible for records created by government and the private sector. Despite this integration, however, the Government Archives Division has recently created the position of Electronic Records Officer to provide expertise in this area. In general, it appears that administrative separation of responsibility for electronic records often results in the separation of descriptive information about them from descriptive information about records in other media created by the same body.

In terms of the original purpose of this investigation, the results of this analysis proved to be singularly unhelpful. It was not possible to find enough examples of "real" fully-processed records to simply "translate" them into *RAD*. The samples of described records are of a very limited range, representing mainly numerical data files rather than records from more complex electronic systems. The handful of records that are more challenging (e.g., those from office systems) provide a starting point for the construction of a *RAD* description, but the contextual relationship with other records in the fonds is not evident. The search for common data elements in the descriptions proved largely meaningless--the obvious ones (title, dates, extent, etc.) are there, but the existing practices are not rigorous enough to determine the value of elements (such as documentation and technical information) that are not used consistently. Unfortunately, the exercise of examining current practices provided few clear answers or solutions. The exercise did, however, provide considerable insight into the challenges of describing electronic records; it also suggested a number of areas that need to be investigated systematically, and questions that need to be answered.

The fact that description is clearly an ongoing process, regardless of the medium of the records, has been widely recognized.⁵ Without sophisticated automated systems, however, archivists have lacked the means to link the functions so that information flows through. Furthermore, as long as archivists are dealing with material that can be understood through visual examination, it is relatively easy (albeit inefficient) to capture it repeatedly. The foregoing examination of current practice provides indisputable evidence that much descriptive information is collected very early in the archival acquisition process, often within the records management functions. For electronic records, which cannot be understood through visual examination, it is essential to capture descriptive information as early as possible⁶ in the acquisition process (at the scheduling/appraisal stage), and to bring it forward as processing takes place. Despite the uncertainty about the best ways to describe electronic records, much descriptive information is captured at the scheduling, appraisal, and accessioning stages. If such a vast store of descriptive information is available, why are there not more "complete" descriptions of electronic records?

There are a number of reasons why so few electronic records have been described. One explanation is the lack of resources affecting all archives: every institution has a backlog, and electronic records are part of that backlog. When resources are scarce,

choices must be made, and it appears that institutional resources for electronic records are going into the front end of archival work (appraisal and acquisition). There are, however, a number of more fundamental issues to be addressed before descriptive problems can be solved. This study raises specific questions that need to be tested in a number of areas: archival theory and practice; the nature of electronic records; descriptive systems, standards, and products; and the administrative context in which electronic records are dealt with.

In the first place, archivists have been ambivalent about how electronic records are treated. Are they (for the most part) simply invisible textual records or are they another medium? They have been treated administratively and within library and archival cataloguing rules as a special medium (akin to maps or photographs), and have not been fully integrated into the descriptive practices and systems of the institutions surveyed. This is a question requiring examination. Closely related is the question of the appropriate administrative context to deal with archival electronic records: are they more effectively managed with dedicated resources, or when they are integrated with other functions? The institutions surveyed in this study provide a body of experience on which to base a comparison of the benefits and disadvantages of integration and separation.

The application of the concept of levels of arrangement to electronic records has not been critically examined. Much of what has been described so far has been arbitrarily designated a "series"--whether it is a system, a file (in the automation sense), or a database. In other cases (such as NAC), data files have not been assigned a level of arrangement at all; they are included in their media-specific section of the inventory for the relevant record group. Linked to the question of levels is the need to establish the relationship of electronic records with other records in other media (or with other electronic records) produced by the same creator. This is rarely done, due in part to the size of the institutions and the separation of the units responsible for electronic records, and in part to the fact that electronic records are rarely part of a closed fonds or record group with up-to-date inventories. Only in the published guides are electronic records set in the context of the related records in the record group. The *NAC Guide to the Holdings of the Government Archives Division* (1991) describes both textual and electronic records (as well as microforms) which are part of the same record group (although they do not include cartographic, graphic, or audio-visual materials); the NARA gopher (which admittedly includes only a limited number of records) does, in some cases, place the electronic records in relation to other records of the same creator, e.g., Records of the Presidential Commission of the Space Shuttle Challenger Accident. On the other hand, other NARA files on the gopher describe only very specific data files, e.g., Internal Revenue Service (IRS) Statistics of Income series and do not relate them to the other IRS records (electronic or textual) that NARA holds.

The wide-ranging nature of the records themselves has impaired description. Attention has focused on data files, but has not moved on to records from other types of systems. Although there is an established body of practice and use in the description of flat numeric data files, these descriptive rules are not adequate for the newer forms of electronic records. Trying to shove the description of word processing files into the descriptive format for machine-readable data files results in a very thin description.

Much of the information required for data files (e.g., principal investigator, unit of analysis, population characteristics) is simply not relevant to a word processing file; conversely, there is no place in the data file description for important information about the word processing file (e.g., file naming conventions, search capabilities, etc.). Different descriptive elements and products may be required for electronic records from different types of systems, particularly at the lower levels of description. Before these descriptive products can be defined in detail, however, archivists must come to some agreement about identifying and defining the different types. A number of suggestions have been made, ranging from definition by the broad administrative function that the application supports (e.g., finance, personnel, registration, etc.) to types of systems (e.g., standard databases, integrated corporate databases, and office systems) to functions of the systems themselves (business document systems, transactional systems, and geographic information systems). Any typology must be open-ended, because systems are evolving; eventually, however, records from all types of systems will require archival description.

Related to this is the question of what the records consist of. Is documentation considered to be part of the records? The three institutions surveyed differ on this matter. While SARA files the documentation on the accession file, NARA includes it as part of the records. The NAC generally treats it separately; however, there is a place on the Accession Control Record form to note if the documentation is part of the data file itself.

This raises the question of just what is documentation, and how does it differ from metadata, accompanying material, system description, and finding aids. The draft *RAD* chapter on electronic records uses all these terms, without fully distinguishing one from the other. Archivists have always had difficulty in arriving at a common terminology, and this area is no exception.

What sorts of descriptive products are needed, and how do they relate to existing standards and products? Descriptive products for electronic records have been of two types: at the summary level (i.e., fonds, series, etc.) and a more detailed finding aid (called a documentation package or user guide). Are these adequate? SARA is producing finding aids in the form of user guides, which provide extensive information to assist the user in determining the relevance of particular records for their research, and how to use them. It would be useful to investigate how researchers use the guides, the resources required to produce them, and their application to more complex systems. At the more detailed level of descriptive rules, a number of questions arise. On the one hand, particular information (e.g., physical description) required to describe other forms of records may not be needed for electronic records. On the other hand, essential information about electronic records is not accommodated in descriptive rules for traditional media. One of the ways in which the draft chapter of *RAD* broke new ground is its recognition of the need for description of the systems producing records.

The need for new descriptive products and practices seems clear, and there are a number of specific areas that can usefully be investigated to meet these needs. Who will take on these investigations? Who will define, develop, and test the proposed descriptive products and standards that result? For a number of reasons, everyone is waiting for someone else to do it. The number of archivists sufficiently knowledgeable, interested, and confident to undertake this task is small. Many archivists feel quite

inadequate to deal with electronic records; some are intimidated by descriptive standards as well. The tone of the literature around electronic records has left many archivists less than confident that they are capable of adapting what they already know about description to these new records. Even if one were disposed to become actively involved with the questions, the workload in most archives breeds great reluctance to “waste” time experimenting if it ultimately proves to be wrong, or if someone else is doing it. Many still think that this is an esoteric area that only the “experts” dare venture into, and are content to wait until the “experts” offer a solution.

Unfortunately, at both national archival institutions, the “experts” are also waiting. Both institutions are in the process of developing archival holding systems, and are continuing to use older procedures until the new systems are ready. The NAC staff are also waiting for the final version of the *RAD* chapter containing the rules for the description of electronic records; their wait has been prolonged.

An overview of the situation suggests that a kind of paralysis has set in, with everyone waiting for someone else to take the initiative. The records will not wait, however, and some archivists have decided not to wait either. In fact, a lot of very practical work is being done on a number of different fronts, as working level archivists in many institutions attempt to deal with increasing volumes of records in electronic form. It is such experimentation (and the sharing of the results) that will lead to the development of a body of practice resulting in workable tested solutions to the new descriptive problems facing the profession.

Various approaches are being tried. A number of archivists in Canada and elsewhere are waiting for Chapter 9 of *RAD* to provide some definitive answers. Even though the final version of the rules will not be available for some time, the draft rules provide a fruitful opportunity to conduct much-needed empirical research by evaluation their usefulness when used to describe archival electronic records. Others perceive metadata to be the solution to all archival descriptive problems. Empirical examination of the potential value of metadata and its relation to cataloguing rules would be a very useful area of investigation. Other possible solutions lie outside the narrow confines of description. As noted, appraisal reports were found to be very rich sources of descriptive information. Institutions could ensure the preparation of detailed written appraisal reports as part of the scheduling process, so that this descriptive information can be carried forward into the next stages of archival processing.

The sharing of information about the experimental work going on in institutions is essential to the building of knowledge. The Association of Canadian Archivists' Special Interest Section on Electronic Records (SISER) has developed an ambitious workplan that emphasizes the importance of communication. Dissemination of information on the evaluation of experiments in description of electronic records could be an important part of this initiative. Similarly, the Society of American Archivists' Electronic Records Section has established an Information Exchange Working Group whose mandate is communication within and outside the archival community.

For those who are less engaged in electronic records issues, another area offers itself as a fruitful area of investigation. Clearly archivists must find ways to capture the information they are already collecting. If archivists are trying to reduce the

effort put into (often repetitive) capture and recording of the same elements of information, they need look no further than their own processes. The potential of modern technology offers a way to improve the efficiency of descriptive practices by capturing relevant descriptive information early in the acquisition process and letting it flow through, enhancing instead of re-doing it. In the shorter term archivists have a far better chance of achieving this, while others try to make sense of metadata. Modern technology also provides the linkages essential to enriching descriptions. Maintaining authority files, or demonstrating the connections between provenance and descriptive information, while clumsy and labour-intensive in a manual system, is now achievable because of rapid advances in technology. The technology that produces electronic records and their inherent challenges is in many ways the very tool that will make archival descriptive systems richer and easier to achieve.

The current literature has focussed on the ways that electronic records are different from more traditional holdings. The fact that there are significant differences, however, does not mean that they must be dealt with in isolation. Looking at some of the similarities allows a more comfortable evolution of practice from the framework of known practice and principles. The archival community has embarked on an exciting examination of descriptive practices, to see where and to what extent electronic records fit in, and the ways in which description will evolve to meet new needs.

Notes

- * In 1994 the author was awarded a Bentley Fellowship at the University of Michigan to investigate aspects of description of electronic records. This article presents the results of her research. The author wishes to thank colleagues at the National Archives of Canada, the National Archives and Records Administration of the United States, and the New York State Archives and Records Administration for their assistance in the preparation of this article. In particular, Yvette Hackett at the National Archives of Canada, Nancy McGovern and Sharon Thibodeau of NARA, and the staff of the Center for Electronic Records at SARA readily provided extensive documentation of their policies and procedures and patiently responded to many follow-up questions.
- 1 See, for example, Margaret Hedstrom, "Descriptive Practices for Electronic Records: Deciding What is Essential and Imagining What is Possible," *Archivaria* 36 (Autumn 1993), pp. 53-63. David Bearman and Margaret Hedstrom, "Reinventing Archives for Electronic Records: Alternative Service Delivery Options" in Margaret Hedstrom, ed., *Electronic Records Management Program Strategies, Archives and Museum Informatics Technical Report* 18 (Pittsburgh, 1993) pp. 82-98.
- 2 At the date of writing (April 1995), draft rules for the description of electronic records had been circulated to the archival community for comments, and it was expected that the approved rules would be published in late 1995. In October 1995, however, the Planning Committee on on Descriptive Standards decided that Chapter 9 of RAD would not be published at this time, as more research is required into the description of electronic records.
- 3 That is, one which describes the records at a collective or summary level--such as the fonds (or record group) or series--as opposed to the file or item level.
- 4 Data structure standards define the elements of information which comprise an information system (e.g., name of individual); data content standards are the rules which define how information is to be entered into the various elements of the data structure (e.g., enter family name first followed by first initial). For further discussion see "Report of the Working Group on Standards for Archival Description," *American Archivist* 52 (Fall 1989), p. 454.
- 5 See, for example, Glen Isaac and Derek Reimer, "Right from the Start: Developing Predescriptive Standards at the British Columbia Archives and Records Service," *Archivaria* 35 (Spring 1993), pp. 86-87; Sharon Gibbs Thibodeau, "Archival Arrangement and Description" in James Gregory Bradsher, ed., *Managing Archives and Archival Institutions* (Chicago, 1988) p. 76.
- 6 The possibility of capturing archival descriptive information by involving archivists in the design of systems is a different question.

SUMMARY OF PRODUCTS DESCRIBING ELECTRONIC RECORDS AT NARA, SARA, AND NAC

DESCRIPTIVE PRODUCT	NEW YORK STATE ARCHIVES AND RECORDS ADMINISTRATION (SARA)	NATIONAL ARCHIVES AND RECORDS ADMINISTRATION (NARA)	NATIONAL ARCHIVES OF CANADA (NAC)
Records management inventory/schedule	Records Series Inventory Form & Instructions	- Sample of "GAPS" database of electronic records scheduled for transfer to NARA, 1993 - Samples from NAPA survey of federal gov't databases, [1990-1] - booklet (<i>Managing Electronic Records</i> , 1990)	Records Disposition Submission Report & Records Disposition Authority (forms and procedures)
Appraisal Report	Sample reports, 1991	- Draft "CER Appraisal Guidelines", 1994 - Sample report, 1990	Archival Appraisal Report (form & procedures) ¹
Accessioning procedure	"Accessioning Electronic Records: An Interim Procedural Manual and Guidelines," 1992 (includes tracking and physical control of tapes)	- Procedure for transfer of electronic records to NARA, 1994 - CER "Accessioning Procedures," 1993 - "Preservation Request for Electronic Records," 1993 - Archival Electronic Records Inspection and Control System (AERIC) guidelines, users' manual and data elements, 1993 - Archival Preservation System (APS) tables and data elements	- Interim procedure paper "Electronic Records Accessioning Procedure" & forms, 1992 - general "Accession Procedure," 1985 - examples of PREL (Preliminary) descriptions - TAPE database data elements
Descriptive rules/ standards	Inhouse rules based mainly on APPM, AACR2, & MARC(AMC) - Title Entry Form - Series Description Form	In-house rules derived mainly from APPM, MARC(AMC), and special formats for items. Rules for Format X (bibliographic record), 1991 - Format X adjusted to accommodate electronic records, 1993	- Machine-Readable Data File Description Form (1980) and data elements (1986) ²
Descriptive record	- Examples of <i>series</i> from in-house database - examples of <i>series</i> from RLIN database	Examples of format X descriptions of <i>series</i> , 1990-93	Examples of descriptions of processed <i>data files</i>
Finding aid	- Procedure manual - Sample user guides	Documentation packages produced but no example available	Samples of documentation packages, 1978-80
Published guide	<i>Guide to Records in the New York State Archives</i> , 1993	<i>A Preliminary & Partial Listing of the Data Files in the NARA</i> , 1994 GOPHER.NARA.GOV	<i>Government Archives Division</i> , 1991 (General Guide Series)

¹ In "Report & Recommendations of the Work Group on Requirements of Submissions, Archival Appraisal & Authorities," Approved 1990.

² Each data element definition includes, where relevant, a reference to the MARC & MARC(AMC) tags, the AACR2 rule, and the SAA data dictionary term.