Will Metadata Replace Archival Description: A Commentary

by **WENDY DUFF***

Résumé

Cet article compare et oppose les principaux concepts de méta-information et de description archivistique écrits par Heather MacNeil et David Wallace. Il met en lumière les similitudes des deux articles et examine leurs divergences. L'article conclut en soutenant que les archivistes manquent présentement de connaissances suffisantes pour répondre aux questions fondamentales concernant le rôle de la méta-information et de la description archivistique formelle. L'article souligne la nécessité d'un programme de recherches qui étudierait les besoins de la clientèle des archives et identifierait les procédés qui protégeraient l'intégrité et l'impartialité des documents et assurerait la saisie de l'information contextuelle pertinente.

Abstract

Heather MacNeil and David Wallace have provided two cogent and insightful discussions on the pros and cons of metadata replacing archival description. On the surface, the two papers seem to be advocating two opposing points of view. Heather MacNeil suggests that archival description should be performed by archivists after records have outlived their usefulness to their creator. David Wallace posits that description at the end of the life cycle causes backlogs, and the loss of vital contextual information. To solve these problems he recommends that creators or systems generate descriptions during records creation and use or what has been traditionally called the active stage of the life cycle. These two papers provide an excellent opportunity to explore this vital issue.

Systems analysts suggest that investigations or evaluations of systems should commence with an examination of any conflicting ideas or alternative points of view. In reviewing the alternatives, one should carefully study any similarities because the converging concepts reveal the essential elements of any system.

These two papers have many areas of confluence. Both MacNeil and Wallace view records as being evidence of transactions, and both assert that description must be context-oriented. The primacy of context, and the need for electronic systems to protect the context, content, and structure of records goes unchallenged. Both speakers reject descriptive practices that disregard the transactional nature of records. Wallace refers to Richard Cox's article on the evolution of the record to support his analysis of records as evidence of transactions. MacNeil, however, notes that the importance of records as evidence of actions has "long been the underlying assumption of descriptive standards and practices directed toward more traditional archival material." Both would agree that to preserve the value of records as evidence of transactions, archival description must capture and link a description of a record to its provenance or context.

Unfortunately, information systems often do not create records, concentrating instead on the preservation of information to the detriment of record-keeping. Concern over this issue has lead Wallace to promote a new role for archivists, one that places them at the conception of the life cycle, establishing standards for record preservation and management as well as dictating record creation. Demarcation between archivists and records managers disappears in this new paradigm and a new role as auditor, system designer, and regulator begins to emerge. He advocates this stance to overcome the threat that electronic information systems pose to the creation of evidence of transactions. He contends that, if archivists do not intercede, records will not survive. MacNeil promotes a slightly different viewpoint. She also accepts the notion that archivists have a part to play in the design and management of electronic record-keeping systems; quoting Luciana Duranti, however, she contends that archivists should not advise creators on what to create.

Should archivists become involved and make recommendations concerning the creation of records? If they do not, will the records needed for legal and administrative requirements or for reasons of accountability be created? What is their role? If archivists use David Bearman's definition of an electronic record as being "communicated across a physical or logical communication switch,"[1] does the question become one of preserving a record rather than creating it? If archivists intercede to ensure the capture of records, do they threaten their impartiality? For example, if an administrator queries a database, gathers information from a particular view of that database, but does not create a copy of that view, has a record been created? If the administrator subsequently makes an important decision based upon that view, might he need a record of what he saw? If an archivist intercedes to ensure the creation of this record, does he destroy the impartiality of the archives? To what degree should archivists assist organizations in the creation of records and what principles should they follow when interceding? If they recommend the creation of records to meet administrative or legal requirements, that is, to meet the needs of the creator, and refrain from any concerns about the production of historical evidence, can they assume the new role advocated by Wallace without impugning the integrity or impartiality of the records?[2] Wallace would argue that archivists must intervene to ensure the preservation of evidence but MacNeil would assert that such actions would damage the impartiality of the entire archives.

This differing emphasis on the impartiality of records is central to the two disparate opinions on metadata's ability to replace description. These two papers,

however, have many similarities or areas of convergence that must also be examined. The two papers agree on the importance of metadata. Metadata are essential if archivists are to maintain the integrity and authenticity of evidence of actions. MacNeil likens metadata systems to protocol registers and sees metadata itself as evidence, as well as a means of preserving evidence. She acknowledges that metadata preserve documentary context and "therefore constitute a kind of description" of a record-keeping system. Wallace also highlights the importance of metadata, outlining the different ways the term has been used. While both authors agree on what metadata are and their importance in protecting the integrity, authenticity, and preservation of records, they disagree on metadata's potential to replace archival description. MacNeil warns archivists against jeopardizing the value of metadata to primary users by altering them to meet the needs of future or secondary users. As metadata are evidence, she would preserve and protect them by ensuring that they are not shaped or moulded by needs of secondary users. For example, just as archivists would not advocate the collecting of data for potential future genealogists through the current registration of births, they should not advocate altering metadata to meet future descriptive needs. Wallace, however, presents no such concerns. After recounting the findings of the New York State Archives project that existing software does not capture a full compliment of required metadata, he states that "clearly archivists need to identify what types of metadata will best suit their descriptive needs, underscoring the need for the profession to develop strategies and tactics to satisfy these requirements within active software environments."

Should archivists articulate their descriptive needs and have these needs met by active record-keeping systems? Even if creators agreed to adhere to these requirements, an idea that MacNeil questions, should archivists attempt to incorporate their needs into all record-keeping systems to ensure that the one or two per cent that come into archival care not require further description? Do the needs of the creators conflict with the needs of the archives? Can the needs of secondary users be met by the functional requirements essential for preserving the context of the documents? If archivists shape metadata to meet the needs of secondary users, will they endanger the impartiality of the records? If, during the active stage of the life cycle, data are created for secondary users, what effect will this action have on the records? How similar are the descriptive needs of primary and secondary users? The primary user will need a system that describes and retrieves electronic records during their active life. Can metadata systems fill these needs? While MacNeil cautions archivists against shaping record-keeping systems to meet their own needs, she does recognize that metadata may eliminate the need for describing records at lower levels. However, she asserts that metadata will not replace all description.

For MacNeil the purpose of description is to preserve, perpetuate, and authenticate meaning over time to ensure its availability to all users. Questions arise over the consistency of descriptions representing different types of material. If metadata could adequately describe record-keeping systems, would the description of these systems be dramatically different from descriptions of a person's fonds or of different types of media? Would these differences confuse users? Would various descriptions impede access to records unnecessarily, accentuating differences

among fonds rather than diminishing them? *Toward Descriptive Standards*[3] recommended that textual records, whether archives or manuscripts, be described using one set of rules. Should electronic records not be described in the same way? Will electronic records created by individuals carry the same data as those generated by corporations or government agencies? Archivists need to understand the impact that different descriptive systems have on users before pursuing the course encouraged by metadata adherents. Empirical research that investigates the descriptive needs of users and answers some basic questions is vital to enable archivists to design better systems and to assist creators in the identification of metadata requirements.

At present, archivists make suggestions or recommendations for systems design based on supposition and conjecture. For example, Wallace states that "archivists have adopted broad collective descriptions of records not because we necessarily wanted to but rather because we had to." Is this true? *Toward Descriptive Standards* recommended that description proceed from the general to the specific. Was this recommendation made out of expediency, or because users must comprehend the totality of the fonds to better understand individual records or files? Do users want broad descriptions or would they prefer item level access without many levels of description? Do they need the broad context provided by fonds and series level descriptions, or will metadata systems that link content, context, and structure obviate the need for higher level description? I would guess no, but I have no data to prove my assertion. Only by undertaking research will archivists move beyond intuition and educated guesses to be able answer these questions.

Research is required to investigate all of the various components of our descriptive processes. To return to the papers at hand, MacNeil and Wallace identified different causes for the backlog of arrangement and descriptive tasks that plague our archives. While they both agree that archivists must ensure that records are adequately managed throughout their life cycle, Wallace asserts that traditional archival description created by archivists at the terminal stage of the life cycle is doomed to failure. MacNeil claims that it is not. Currently, almost all archives face huge backlogs in the processing and description of records. Is this backlog due to [illegible] or appraisal techniques compounded by uncoordinated descriptive systems? If archivists integrated their arrangement and descriptive systems with records management, would the backlogs disappear? Can archivists develop systems that build upon record management and utilize their descriptions in archival systems? Will archives use *Rules for Archival Description* (*RAD*)[4] to create fonds and series level descriptions, relying on record management or metadata to provide control and description at the file and item level? Will this strategy enable archivists to describe their materials within existing resources?

Will metadata replace description? Before attempting to answer the question I shall first consider the findings of two interesting studies: one conducted by information scientists and the other by an anthropologist. They both provide interesting insights into the issue at hand.

In the 1960s and 70s the development of full-text retrieval systems was seen by many as the solution to libraries' retrieval problems. Many researchers believed that full-text retrieval systems would access documents directly, eliminating the need for expensive manual indexing. They predicted the demise of the professional

indexer and heralded the abilities of the computer. The prophesies bear interesting similarity to the current claims of our metadata gurus. In 1985, Blair and Maron,[5] two information scientists who questioned the performance claims of full-text retrieval systems, conducted an experiment to test the assertion that full-text retrieval provided satisfactory access to material. They used a legal database containing approximately 350,000 pages of text and an information request provided by lawyers working on a court case. This situation reflects real needs and followed established procedures. To almost everyone's surprise, except perhaps that of the investigators, the experiment revealed that retrieval was well below satisfactory levels. The lawyers had stipulated that they required high recall or seventy-eight per cent of all relevant documents. Using various techniques the searchers achieved only a twenty per cent recall rate; in other words, they did not retrieve eighty per cent of the relevant documents in the database. Since this experiment, information scientists have sought to improve the recall and precision of full-text retrieval systems. They generally acknowledge, however, that the best results are realized on systems that supplement full-text retrieval with controlled indexing terms provided by professional indexers. The full-text retrieval debate continues in information science circles. Wallace contends that systems can produce metadata that would eliminate the need for archival description. He may be right, but he will need to prove it. Build a system that uses metadata to meet all the descriptive needs of primary and secondary users. While the project being conducted at the University of Pittsburgh, described by Wallace, moves the profession in the right direction, much work remains to be done. Before archivists abandon archival description, they require research that compares the retrieval performance of the two types of systems: one containing descriptions consisting of metadata and the other with descriptions supplied by archivists.

Another study that may be more apocryphal than true, but is relevant to the present discussion, involves a tribe of Brazilian natives.[6] When studying the hunting behaviour of this tribe, anthropologists noticed that at the beginning of each hunt the hunters would run 250 yards out into the forest, turn left, and then run another 100 yards. They would then jump or leap into the air as though crossing a barrier and then continue on toward the hunting ground. No logical reason could be discovered for this ritual until one anthropologist discovered that, many years before, a river had existed about 250 yards from the village and that the river narrowed at one particular point. The natives had always run to the river, travelled to where it narrowed, and then leaped across it. Although the need for this behaviour had long disappeared with the drying up of the river, hunters continued to begin their hunt following established rituals.

Will the need for archival description similarly disappear? Or will archivists continue to need to describe material as MacNeil asserts, to "communicate knowledge about the broad administrative and documentary context"? Will metadata requirements fulfil the descriptive needs of an archives' secondary users or will metadata require supplementary descriptions? I do not know. To answer these questions archivists need to conduct research projects. How well are archivists presently fulfilling user needs? How well will metadata systems fulfill them? Will interfering in the design of metadata systems threaten the impartiality of the records? What metadata can a system generate after the records become archival? Can archival

systems scan the records, identifying types of records by their form—perhaps conducting a diplomatic analysis of the record-keeping systems? How many of the elements required for description could the system supply? Is a description of record-keeping systems sufficient? MacNeil points out that organizations have both electronic and non-electronic record-keeping systems. How will archivists accommodate these other types of systems if they concentrate all of their efforts on developing metadata systems? Do users require description of all of the record-keeping systems of a creator as *RAD* would assert, or would separate and perhaps incompatible descriptions of these systems suffice? If research discovers that metadata systems provide sufficient retrieval, archivists should stop describing records and redirect their energies. Creating descriptions that are not needed is as misdirected as following rituals to get over a non-existent river. However, archivists must first study their user needs, identify processes that protect the integrity and impartiality of records, and ensure the capture of important contextual information.

The first stage of a research project identifies the important components of the situation being studied. Heather MacNeil and David Wallace have identified the important concepts, delineated the problems, and provided potential solutions. These two papers provide an excellent analysis of two different strategies. Now archivists must gather empirical data to guide their future actions as they strive to preserve and make available the evidence of past and future transactions.

Notes

* The original version of this article was a commentary on the papers by Heather MacNeil and David Wallace presented at the annual conference of the Association of Canadian Archivists, Ottawa, 25 May 1994. The articles based on these two papers are printed immediately before this commentary in the current issue of *Archivaria*.

1 David Bearman, "Archival Data Management to Achieve Organizational Accountability for Electronic Records," *Archives and Manuscripts* 21 (1993), p. 17.

2 Jenkinson recommended that Administrator bodies should destroy what they no longer need provided that the Administration proceeds from "the needs of its own practical business; provided that is, that it can refrain from thinking of itself as a body producing historical evidence. Sir Hilary Jenkinson, *A Manual of Archive Administration*, rev. 2nd ed. (London, 1966), pp. 149-150.

3 *Bureau of Canadian Archivists, Toward Descriptive Standards, Report and recommendations of the Canadian working group on archival descriptive standards* (Ottawa, 1985).

4 Bureau of Canadian Archivists, Planning Committee on Descriptive Standards, *Rules for Archival Description* (Ottawa, 1990).

5 David C. Blair and M.E. Maron, "An Evaluation of Retrieval Effectiveness for a Full-text Document-retrieval System," *Communications of the ACM* 28 (1985), pp. 280-299.

6 This story was told to me by an anthropologist, but unfortunately I cannot located a published source to confirm it.