

Authority Control: Beyond a Bowl of Alphabet Soup

by CYNTHIA J. DURANCE*

This paper is a general overview of the concept of authority control and of the reasons why authority control is an important tool for the achievement of intellectual access to archival materials. When I chose the subtitle (long before I wrote the paper), I was reminded of my childhood, when for lunch we sometimes were served a well-known brand of soup which contains pasta shaped like the letters of the alphabet. We would always search the soup and carefully extract the letters of our names and arrange them on the plate. This urge to create order out of chaos and meaning out of the contents of disorder begins when one is young. Providing authority control is not unlike the process of creating order and meaning out of a bowl of alphabet soup. Authority control provides the means to create intellectual order out of the chaos of documents received in an archival repository. It functions like the road-map that guides one through the intersections of complex freeways and allows one to choose the best route towards a particular destination. The road-map is a surrogate for the actual roads; likewise, authority control is a map which guides the user through the maze of archival holdings towards the documents they seek. The user of authority control systems is like the child who, when asked whether he knew how to get from one city to another, replied, "You get in the car and follow the red line."

What Is Authority Control?

First of all, there is the need to define terms. *Authority control* is an umbrella term for the means of providing access to the holdings of a repository through names, subjects and functions — these being constructed in such a way that their relationships, or non-relationships, to each other are apparent. In effect, authority control in an archival setting introduces order into, and a road-map through, the complexities of corporate organizations and their predecessors, personal names, functions, and subjects of documents, and their interrelationships. It also provides provenancial data and documentation about material held in the archival repository(ies).¹

Authority control has three components: the first is an intellectual component, while the other two components are physical:

- 1) *Authority work* is the intellectual exercise by highly skilled professionals in order to research and establish the authoritative access points, their scope and relationships to other access points. It is the most labour-intensive aspect of authority control. Authority work consists, first, of “determining that an [intellectual] relationship exists [among access points]; secondly, establishing and linking all the possible access points which could reasonably express [that relationship].”²
- 2) An *authority system* is the application of sophisticated techniques (preferably automated) to record, maintain, manipulate and display relationships among related access points. True automated authority systems are “linked” or “interactive” systems which record authority data separately from the descriptive records of documents to which they pertain, but which connect the records in such a way as to make this separation transparent to the user. This capability is even more important for archives than for libraries, because the ability to link the history of documents with the history of their creators is central to archival description.
- 3) An *authority file* is the database created by the input of the data which results from the authority work into the authority system. If the system is automated, this data can be variously selected, sorted and manipulated by software programmes and organized on the screen or in print in various formats, or in what computer scientists would call various “user views.” If the system is manual, then user access paths are necessarily more limited and less amenable to data manipulation.

The database has a dual function. It is neutral in the sense that it does not describe documents *per se* but rather the relationships between documents and their creators, and in that it describes intellectual, not physical, relationships. Therefore, the database can stand by itself as a complete listing of names of individuals and corporate bodies, as well as of their administrative histories or biographical sketches; of functions; of documents and their current and historical relationships to each other; and of the interrelationships of the subjects of documents. The database also contains the pointers necessary for its records to be combined by means of software with the physical descriptions of fonds, series and files, thus providing a fully integrated access tool.³

Speaking of the structure and content of archival description, Heather MacNeil has written,

Although archivists have not traditionally distinguished between them, the description of a fonds and its parts and the provision of access points to those descriptions (that is, the indexing of descriptions) are separate, albeit intimately linked activities and should be treated as such.⁴

In order fully and economically to reap the benefits of authority control, it is necessary to separate logically the information about the organizations which at various times controlled the creation of the documents from the information about the archival documents themselves. The two must be rigorously separated in authority work so that the links from one to the other can be unambiguous as they mix and match over time. In this way, the authority system can provide the essential links between the provenance of the records — their attachment to a fonds — and their description, physical arrangement and custodial history. This is a subtle but crucial distinction, as the different elements seldom run parallel over time; it is perilous, though easy, to confuse them.

Descriptive information (i.e., Part I of the *Rules for Archival Description [RAD]*⁵) can vary to a certain degree from the content data standard without serious harm to the database (although at some cost to the accuracy and flexibility of retrieval). However, authority information for access points (i.e., Part II of *RAD*) must always be consistent and up to date if the file is to be adequately searchable. Without a rigorous structure there is no way of linking similar material for which access points have been created in different parts of the organization or indeed of the country.

When one contemplates some of the large and constantly changing departments of government, one can understand the rigour with which authority research must be done and the links made and maintained in order to connect large, complex and changeable organizations with equally large, complicated and changeable series of documents. Furthermore, in the case of the National Archives of Canada, this must be done in two languages and the links among the parallel headings and references maintained.

In a manual environment, combining provenance information with descriptions in a single narrative is possible because the user can make certain intuitive connections between the two. However, these connections can only be made in a linear dimension and are far from optimal. In the computer environment, however, combining these two components of provenancial and documentary histories into one record does not work to the user's advantage. The computer being systematic but a totally logical idiot always taking logic to its most absurd conclusions, we humans must impose on it the intricacies and nuances which we wish to have considered. Nevertheless, if rigorous logic is used for segmenting the data input, one of the benefits of computers is that software can be built to display records in different combinations — hierarchical, relational, etc. In other words, the authority records can reveal to a user information about relationships between records and their creators — the fonds to which they belong — or can also be combined in different ways to reveal information about the creators or the documents themselves. (It is important to emphasize the importance in an automated environment of not confusing the data to be input with the data to be presented on the screen in response to a specific enquiry. Requirements for data input may seem rigorous and excessive to some, but the rigour of input has a direct relation to the flexibility of output. The old adage of "garbage in, garbage out" is unfortunately only too true and largely immutable, as many archivists have found, to their sorrow.) The cost of research and of the input of data according to consistent standards, allowing the description of highly complex links, will not be trivial and argues clearly for a national, cooperative approach where the work, as well as the benefits, are shared.

What Kinds of Authorities Are There?

In the archival application, authorities may exist for

- 1) corporate names, their histories and functions (including names of government organizations, families and corporations);
- 2) personal names (individuals) and biographical details about them;
- 3) subjects (including names used as subjects and geographic names); and
- 4) functions, occupations and/or forms of material.

All these types of information require a controlled vocabulary within the authority file. Without a controlled vocabulary, one cannot ensure the retrievability of the records sought and *only* those records. There is no coherent road-map. By “controlled vocabulary” is meant the totality of all the identified access points, the preferred terms and the linkages established among preferred and variant terms within a particular file. The linkages can be of three kinds:

- 1) An equivalence linkage connecting synonyms. For example,
United Nations Educational, Scientific and Cultural Organization
for works by and about this body, search under Unesco
- 2) A hierarchical relationship linking broader or narrower terms. For example,
STREETCARS
for related materials, search also under PUBLIC TRANSIT VEHICLES
- 3) An associative relationship linking terms in some significant way, including other language links. For example,
BUSES
for related materials, search also under STREETCARS

or

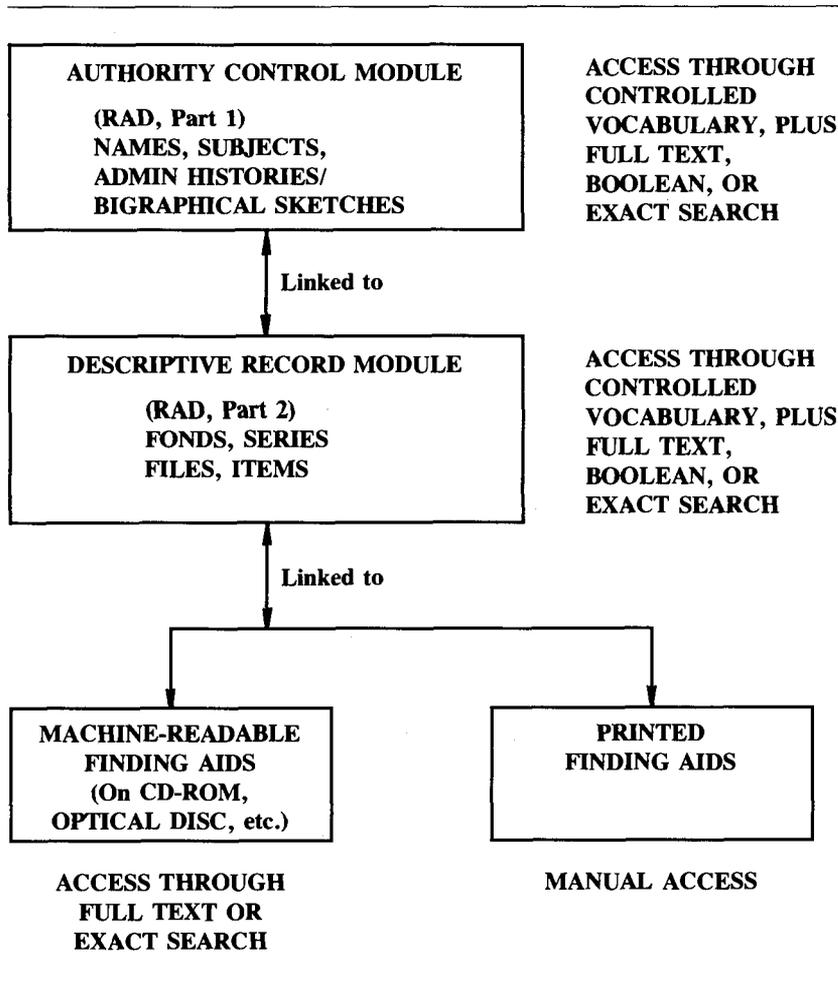
Canadian Council of Archives

*for related materials, search also under Conseil canadien des archives.*⁶

The reason why a controlled vocabulary is needed is to allow the user the cheapest and most efficient path of accessing the data. Furthermore, and importantly, a controlled vocabulary allows searches through many parameters and combinations of parameters —physical format, chronology, scale (for cartographic materials), etc., because the data input is very precisely coded.⁷

Some experts maintain that full-text retrieval of free text obviates the need for a controlled vocabulary. However, a full-text search on unstructured text does not provide the context (so essential to archival work) in which the retrieved word or phrase is used. For example, a full-text search of free text on a word such as “painting” will retrieve everything from works of art to house painting, whereas a similar full-text search on a controlled vocabulary database will distinguish among the different uses of the word. In a database the *raison d’être* of which is to help the searcher to navigate through the intricacies of names or subjects and their definitive relationships, full-text retrieval imposed on a controlled vocabulary database is the best combination, ensuring that one retrieves everything which is wanted and nothing which is irrelevant. A controlled vocabulary database provides the structure of a high relevancy/low recall mechanism. Full-text search on free or unstructured text, however, is a high recall/low relevancy mechanism. Full-text retrieval on free or semi-structured text can be used to good effect in archives for searching detailed finding aids on electronic or optical databases. To return to the road-map analogy, when one arrives at the city of one’s destination, it is usually most efficient to switch to the city map for guidance within a known matrix, i.e., the city, to one’s precise destination. **Figure 1** illustrates schematically the relationship among the different aspects of an archival intellectual control record — authority control, description and detailed finding aids.

Figure 1: Schematic diagram of an archival holdings system.



The library world, which pioneered authority control, has found that personal names are the least labour-intensive to control, while subject headings are the most labour-intensive. Corporate names, which are a major preoccupation of archives, have traditionally fallen somewhere in the middle. However, as archivists try to chart the very complicated organizational histories of corporate creators and link them to different, but equally complicated series of documents, one wonders whether corporate names will not prove even more difficult to control than subject headings. The number and complexity of the links required among creating agencies and between records series and files, as well as among themselves, are somewhat daunting. When one takes into account the complexities of linking all of them in two languages, then the task is truly a challenge — one which, however, if it can be met, will provide enormous benefits to both archivists and users.

What Are the Benefits of Authority Control Which Make It Worth the Investment?

1) *Economic*

First and foremost the benefits are economic; the benefits are long-range, however, while the initial costs of data creation and input are high — a formula which is not easy to sell to any but the most enlightened administrators. Nevertheless, once created, the same authority data can be used for many different descriptive records and hence will exponentially reduce expenditure. “Do it right once and use it many times” is the rationale. Linked authority control systems allow for what is called a “global update”; that is, a change in the name or structure of an organization or a series of documents can, through one computer command, be reflected in all the records in the database which are linked to that authority record. Thus the savings in staff time needed to update and maintain the integrity of the database are enormous. These savings become increasingly necessary as archivists respond to the ever-accelerating pace of change in societal and governmental administrative structures and functions. Creating access points and their linkages accounts for about 75 per cent of the cost of providing access to information, whereas the descriptive portion accounts for the other 25 per cent. To create and re-create this structured authority information in different ways in different parts of an organization or country does not make economic sense. To create it authoritatively once and for everyone, however, does make economic sense, as does sharing the cost of its creation.

2) *Intellectual Integrity*

The accurate and consistent application of the core archival principles of provenance and *respect des fonds* is now possible in a way which was not feasible under disparate, individual conventions within and among repositories. Using authority control, the principles of provenance and *respect des fonds* can be maintained intellectually, even though, for various practical and conservation reasons, parts of the fonds or series (especially those in different media) are physically segregated and maintained in separate, often remote, locations. This fact has particular importance at a time when holdings of many repositories are scattered among different physical locations and when government policy in Canada is proposing that national cultural heritage should become increasingly decentralized. The result may be that the parts of a fonds are physically dispersed among many different repositories and/or locations.

3) *Retrieval and Service to the User*

The existence of properly constructed and maintained authority control, through its automatic connections to all documents pertaining to a particular corporate body, allows interactive computer searching whereby the users can define and refine the search. (This statement, of course, assumes that there is a sophisticated and complex automated authority system in place and that the data input has been correctly coded and linked.) By “users” is meant not only the researcher and the general public who visit the archival repository, but also the archivist whose job it is to serve them. To paraphrase Ronald Hagler, a large part of the total cost of access control is the user’s own time spent searching. If a research tool is well-constructed, then the user finds the information efficiently and with less dependence on help from the repository’s staff. If the tools or technology are inadequate, however, then the user wastes time and must also receive more help. For most archival users, as well as for archival staff, time is money; archivists must therefore seek to serve everyone as effectively and efficiently as possible.⁸

4) *Sharing*

The ultimate result of good, consistent and up-to-date access to the holdings of a repository is the ability to share it with other repositories, and hence with their users. This must be a major public-service goal if archivists wish not only to preserve but also to make known the memory of the nation. Without authority control, sharing of data and access to information is only possible in the most labour-intensive and ad hoc manner.

Requirements for Authority Control, or How to Get There from Here

It is beyond the scope of this paper to lay out a detailed blueprint of the steps necessary to achieve authority control in archives on an institutional, national or international level. However, the paper can perhaps identify some general requirements and issues to be addressed, and suggest useful initiatives to be undertaken.

1) *Senior Management Support*

The ongoing commitment by the archival profession and by the senior management of archival institutions to the implementation of authority control is essential, as is the attendant need to reallocate sufficient resources to achieve the consistent creation of archival data and the implementation of standards format. Let us not delude ourselves; these resources are substantial. They are, however, already being expended in archival repositories, but in less concentrated, productive and visible ways. It is therefore a question of clearly identifying and targeting the most cost-effective methods. No one will be able to reap the numerous benefits of authority control unless everyone understands and accepts its inherent discipline.

2) *Organizational Adjustments*

Organizational adjustments will, in all probability, need to be made in order to ensure the consistent implementation and quality control of the data and format standards throughout corporate databases. This may involve an investigation of the desirability and efficiency of functional specialization by archivists, and also the creation of central maintenance for the data files as a whole.

3) *Education*

Assuming that greater separation of functions is required, training must follow suit, and that training of staff will involve resources. More archivists need to become expert in control theory and practice and in data and format standards and also to accept this role as a creditable and essential part of their professional culture. This, however, will not be achieved without concerted effort.

4) *Software and Data*

An investment in the acquisition of complex computer software is essential, as is the investment in its maintenance and upgrading. It behooves archives seriously to investigate acquiring "off-the-shelf" existing authority control systems, thus avoiding a major investment in time and money. These are very complex systems (particularly if bilingual), which have already been developed and implemented by libraries, and more recently inaugurated by museums, at considerable cost in public funds. Libraries have had well-designed automated systems for twenty years — so well-designed, in fact, that they have remained practically unchanged since their inception. They work well for libraries. With full awareness of the differences between archival and library

applications, therefore, let us test to see whether these systems can also work for archives and where they may fall short of our requirements.

Any existing system which satisfies 60-75 per cent of archival needs has a better potential for achieving a high satisfaction rate than any system built from the beginning will have over the first three to five years of its life. Wherever possible, archivists should want to use data created by others because, instead of reinventing the wheel, it is more sensible and productive to concentrate energy and scarce resources on inventing a better tire. Applications of this strategy are the large name and subject authority databases already built in North America by the National Library of Canada and the Library of Congress (at great cost yet also great benefit to both governments). In fact, these files existed in manual form much earlier. The large bibliographic utilities, such as UTLAS, RLIN and OCLC, and many other autonomous library systems, also have linked authority systems, using these national authority files as their basic input.

By working cooperatively with these institutions, it is possible to enrich the authority databases of archives, museums and libraries. Libraries, faithful to their mandates, only record the organizations or sub-organizations which have published in their own right. They also record name changes, however, and abbreviated administrative histories based on limited research. Hence there is much that archives could do to adapt and enrich the existing databases both in depth and breadth of coverage. The same is true for personal names given that many prominent citizens are not published authors and are therefore not included in library authority files unless their name is used as a subject heading.

Conclusion

There are four major steps needed to achieve authority control in archives:

- 1) management commitment
- 2) reallocation and/or specialization of existing resources
- 3) education and training
- 4) commitment to use appropriate existing software and data wherever it can be found

What a boon it would be to researchers and archivists alike to be able to access a system and locate the holdings of all the repositories in Canada or, better still, in North America. What if, for example, one could know in detail where all the documents were located which pertain to Cornelius Van Horne, the railway magnate, or to the St. Lawrence Seaway and its impact on international trade in the twentieth century? Furthermore, what if one were able to access the provenancial context in which all these documents were created and understand their links to each other? If information is both power and a commodity which has intrinsic and monetary value (as all our policy-makers have been saying for two decades), then we archivists would have created a significant national asset.

Notes

- * Paper presented at the 1992 Annual Conference of the Association of Canadian Archivists, Montréal, 12 September 1992; revised and edited for publication.

- 1 Max Evans has expressed it as “authority control ... an intellectual construct, susceptible to emendation as institutions evolve, as functions change, and as administrative structures are altered”: Max Evans, “Authority Control: An Alternative to the Record Group,” *American Archivist* 49 (Summer 1986), p. 1.
- 2 “Access Point: A name, term, etc., under which a descriptive record may be searched and identified”: *Rules for Archival Description* [Glossary] (Ottawa, 1990), p. D-1.
- 3 Ronald Hagler, *Bibliographic Control and Information Technology*, 2nd ed. (Chicago/Ottawa, 1991), p. 76.
- 4 Heather MacNeil, “The Context Is All: Describing a Fonds and Its Parts in Accordance with the *Rules for Archival Description*,” *The Archival Fonds: From Theory to Practice/Le Fonds d’archives: de la théorie à la pratique* (Ottawa, 1992), p. 217.
- 5 *Rules for Archival Description* (Ottawa, 1990).
- 6 Hagler, *Bibliographic Control*, p. 77.
- 7 For more detailed information on types of authorities, their structure and maintenance, see Elizabeth Black, *Authority Control: A Manual for Archivists/Le Contrôle d’authorité: un manuel destiné aux archivistes* (Ottawa, 1991).
- 8 Hagler, *Bibliographic Control*, pp. 23-24.