The case of Tweddie-dee and Tweddle-dum: Authority control in a retrospective database[*]
La similitude des gouttes d'eau: le mot décisif dans les fichiers retrospective

By: Clara M. Chu, Judith Green, Bertrum H. MacDonald, and R. Alan Richardson

Chu, Clara M., Green, Judith; MacDonald, Bertrum H.; and Richardson, R. Alan. "The Case of Tweddie-Dee and Tweddle-Dum: Authority Control in a Retrospective Database," *Canadian Journal of Information and Library Science*, 12(1): 35-57, 1987.

Made available courtesy of University of Toronto Press: http://www.utpjournals.com/cjils/cjils.html

***Note: Figures may be missing from this format of the document

Abstract:
Authority control is increasingly becoming an accepted element in the design of online catalogues and databases. Much of the literature on this topic focuses on online catalogues and considerably less attention is given to databases of retrospective literature. In this paper the application of authority control to three aspects of the structure and design of Science and Technology in Canadian History: A Bibliography of Primary History: A Bibliography of Primary Sources to 1914, namely, author-names, subject-descriptors, and serial titles, is described. The creation and structure of each authority ale is described, the attendant problems associated with authority control in such a database are noted, and implications for further work of this: nature are discussed.

Le contrôle d'autorité est un élément qui prend de plus en plus d'importance dans le développement des catalogues en ligne et des bases de données. La littérature sur le sujet met l'accent sur /es catalogues en ligne et accorde considérablement moins d'attention aux bases de données rétrospectives. Cet article décrit l'application du contrôle d'autorité à trois éléments de la structure de Science and Technology in Canadian History: A Bibliography of PriCanadian History: A Bibliography of Primary Sources to 1914, soit les noms d'auteur, les descripteurs et les titres de périodiques. Onles descripteurs et les titres de périodiques. On décrit la création et la structure de chaque fichier d'autorité, on note les problèmes associés au contrôle d'autorité dans une base de données de ce type et on discute de l'intérêt de ce projet pour d'autres travaux du même genre.

Article:
Introduction
Authority control has become an important "buzz" term among those dealing with automated library catalogues. Robert Burger has written that "in spite of its detractors, authority control has come to he recognized as an essential component of any on-line catalog" (Burger, 1985, p. 1). The recent literature is replete with articles and books on the topic (e.g., Auld, 1982; Clark, 1983, 1985-86; Rochefort, 1985).

The concept of authority control is, of course, not a recent idea. Writers earlier in this century, such as Charles Cutter, commented on the topic (Cutter, 1904), but with the advent of online catalogues interest in and work on authority control has been spurred. Early writers on authority control saw its application to the need for prevention of duplication of work; previously authority files have been seen largely as a convenience for the cataloguer. More recently, the purpose of authority control has been described in terms of its support for collocation and syndetic functions of a catalogue or database.

Authority control is the generic term for the activity associated with the creation and maintenance of authority files and the linking of those files to bibliographic records in order to ensure that two purposes of a catalogue

are provided, namely, the finding function and the gathering function. Burger has elaborated on this activity in some detail but uses the phrase authority work rather than authority control (Burger, 1985).

Virtually all of the literature on authority control is focussed on its application to current day online catalogues which by their very nature deal primarily with monographic literature. Almost no attention has been given to applications of authority control to databases containing mostly periodical literature and even less to those of a retrospective nature.

In this paper, we describe the use of authority control in the generation of a bibliography of science and technology in Canada prior to the First World War. In particular, we discuss the design, construction, and maintenance of three authority files and note their function in relation to the bibliographic database. Further, we discuss some of the major difficulties in developing authority files for a retrospective bibliography and set out some implications of authority control in future applications.

First, we need to point out that while authority control is normally associated with the provision of authorized headings or access points for a catalogue or database (e.g., personal-, corporate-, conference-, and place-names; uniform, conventional, and series titles; and subject headings), we have adopted Malinconico's definition that an authority file can only be used to provide effective control over those "elements that can be expected to be common to a number of individual bibliographic records" (Malinconico, 1982, 7). For this Bibliography this was taken to mean that authority control could be applied to not only those elements used as access points, e.g., author names and subject descriptors, but also to serial titles because the latter were a common element in many records of the Bibliography. Serial titles could be used as an access point (we expect that in a potential online version of the database some users might wish to search on serial title) but in the published microfiche version (Richardson and MacDonald, 1987) access by serial title is not provided.

Description of the Bibliography

For the purposes of this paper a brief description of *Science and Technology in Canadian History: Δ Bibliography of Primary Sources to 1914, Canadian History: Δ Bibliography of Primary Sources to 1914,* which has been discussed in more detail elsewhere (MacDonald and Richardson, 1983), is appropriate. While the early work for the Bibliography can be traced back to the mid- 1970s, the current product (Richardson and MacDonald, 1987) is the result of a three-and-a-half year project begun in May 1983. The Bibliography is published in microfiche format and contains over 58,000 unique records of published primary works dealing with science and technology. Secondary works, such as biographies, and historical discussions of scientific disciplines, institutions, and manufacturing firms, have been excluded. Entries were created for works published in Canada or abroad written in English, French, or other languages, with the large majority being in English. Of the 58,109 records in the database, 52,419 entries were created for periodical articles, the remaining 5,690 entries for monographs.

In all cases a copy of the work was seen prior to the preparation of an entry. Research assistants visited libraries in locations all across Canada and in the United Kingdom to obtain the data for the Bibliography. Over 200 Canadian and foreign periodicals were systematically searched and monograph holdings of several libraries were examined to obtain the data for the Bibliography. The data was collected with microcomputers and later transferred to files on a mainframe computer for editing and manipulation using DPL (Data Processing Language), a database management system. Approximately two years were devoted to data collection and one year to design and editing of the database.

Not surprisingly, a number of definitions had to be built into the design of the Bibliography. First, Canadian science and technology was taken to mean any scientific or technological work published in Canada or abroad by a Canadian, or published inside or outside of Canada by a non-Canadian if the work contains some substantial material about Canada. Canada was taken to mean the territory circumscribed by the current national boundaries. This definition largely follows that adopted by the National Library of Canada and the Canadian Institute for Historical Microreproductions. The second definition concerned time parameters. Literature cited in

the Bibliography was published from the mid- sixteenth century to the end of 1913. Most of the entries, however, represent literature of post-1850 imprint.

Since it is anticipated that historians of science and technology will be the primary users of the Bibliography, the final product was created with such an objective. The database was sorted to produce three main catalogues and a number of indexes. All works with known authors are listed alphabetically by author in the author catalogue (multiple works by an author are sorted chronologically by date of publication and within date alphabetically by title). All entries in the database are listed alphabetically by title in the title catalogue (works with the same first fifty characters are sorted chronologically by date of publication), The subject catalogue contains all the entries in the database listed according to a subject scheme developed for the Bibliography. In addition to the three catalogues, a KWOC index on titles was prepared and an index of monograph entries based on the first fifty characters of the titles was generated. As a further aid to users both the serial title authority file and the subject classification scheme are reproduced on microfiche.

This brief description of the Bibliography highlights the range of publications included, in terms of subject, type, and date of publication, and leads to a discussion of the access points provided for the database and the three authority files needed to support them (see Figure 1).

Authority control applied to author names

An author name authority file was created from the beginning of the project because of the need to have author name integrity in the database. Such a subsidiary file was not needed for the pilot project of 3,000 records (MacDonald and Richardson, 1981) but even there different author names were standardized when variations were encountered. After having estimated that there would be well over 10,000 bibliographic entries in the final database and realizing that several research assistants would be involved in the data gathering stage, control had to be established that would direct the research assistant(s) to the authoritative form of a name and that would ultimately provide the mechanism for proper collocation of records in the final author catalogue.

Originally, the name authority file was created for dual purposes, namely, to provide author name authority control and to act as a search strategy tool. The first objective established one form of a particular name as the authority such that its use, coupled with cross-references from its variants, ensured that all works by one author would be filed together under the specific format of the name. The second objective was to provide a list of authors that could be used to search for additional entries for the Bibliography. As the author name authority file grew to several thousand names, however, the file became too cumbersome to use as a searching tool.

The authority file was built first by compiling names from several sources: (a) all the names used in the pilot project; (b) the names in a working list of Canadian scientists provided by Dr. Richard Jarrell of York University (obtained from his own work (Jarrell and Roos, 1984) and that of Philip Enros (Enros, 1985)); and (c) appropriate names extracted from *Who's Who in Science International 1914* (Stephenson, 1914). To this initial list, names were added as they were generated from the bibliographic entries.

When an authority record was created for a name to be added to the name authority file, the following sources, listed in priority order, were searched for an already established name authority:

a. *Canadians Authorities* of the National Library of Canada (NLC) b, *The Notional Union Catalog Pre-1956 Imprints* (NUC)
c.   Catalogue of the Canadian Institute for historical Microreproductions
(CIHM)
d.   British Library *Catalogue* (BL)
e.   *Canadians Monographs 1867-1900* of the National Library of Canada (Retro)
f. *Library of Congress Authorities* (LC)
g. Library Catalogue of the University of Western Ontario (OLU)

The use of online bibliographic utilities, such as UTLAS and OCLC, were not considered because of cost and the fact that they duplicate the records in some of the sources noted above.

While published lists of established name authorities are quite comprehensive for current author names, they are less useful for names of authors of early works, i.e., works published prior to World War I. Also, cataloguing agencies only establish name authorities for authors who published monographs while many of the authors in the Bibliography only wrote periodical articles. Only when an author of a periodical article also published a monograph was the name found, if at all, in the name authority sources noted above.

When an established name authority was not found in the sources noted above, a name authority was created by the staff of the Bibliography. Before creating an authority, i.e., using the name as found on the published work as the authority, biographical sources were consulted (see Appendix A for a list of the biographical publications consulted). When a name could be confirmed in a biographical source, it was entered into an authority record in the form for name authorities (see below). Full first names, if available, were used instead of initials. The biographical source, where a name was confirmed, ωας noted in the author annotation field of the author name authority record. biographical source, where a name was confirmed, ωας noted in the author annotation field of the author name authority record.

The format used for a name authority was: surname, first name(s), qualifier. The authority name was sometimes different from the form of the name that appeared on the publication itself. The qualifier served to distinguish writers with the same name and consisted either of birth and/or death dates of the author, or subject interests of the author, or geographical location of the author. The use of subject and geographic qualifiers ωας avoided as much as possible because the subject and geographic qualifiers ωας avoided as much as possible because the subject interests and place of residence of an author could change. In certain cases, modifications were made to established name authorities to facilitate proper sorting b_y computer. For example, the date qualifier, "b. 1850," was changed to "1850-," and honourific names, e.g., Sir, were transposed from a position preceding to a position following first names. In a number of cases, the established name authorities only contained an author's initials while the name on the title page of a publication gave the full name. In such instances a note ωας made of this fact in the annotation field of the name authority record.

An author name authority record contained five elements or fields as follows:

a) AUTHOR NUMBER, which was a unique number assigned to each name authority record. Records were numbered sequentially as they were added t0 the authority file. The Author Number was entered in the author field of the bibliographic entries in the database to represent the author responsible for th_e work (see Figure 2).

b) AUTHOR NAME. (up to a maximum of fifty characters), as it appeared on AUTHOR NAME. (up to a maximum of fifty characters), as it appeared on an actual work. This field was used to sort the alphabetical listing of th_e authority file.

c) AUTHOR ANNOTATION, which included any information about the author that appeared in the work or in an established authority source which helped to identify the author. Such information usually included the topic(s) and date(s) of the author's publication(s), professional affiliations of the author, the place of residence of the author, the biographical source if one was used to confirm a name, and cross-reference information, initially, the author annotated on field of the authority record was often left blank but, as the name authority file grew to several thousand names, it became necessary to provide more extensive annotations to ensure that the correct authority was created and/or used. For example, when there were two name authorities, "Patterson, John" (even with different birth dates -- 1857 & 1872), without an annotation one could not easily determine which one wrote on astronomy.

d) AUTHORITY NAME, as extracted from an established name authority source, or as modified from an established name authority source, or as created by the staff of the Bibliography using the format noted above.

e) AUTHORITY SOURCE, a numeric code, which translated to the name of the source used to create the name authority. For example, *Canadiana Authori-Canadiana Authorities* was assigned the numeric code 1, and an authority created by the staff of the Bibliography was assigned the code 5.

Names were added to the name authority file prompted b$_y$ new names encountered in the bibliographic transcription work. A small working file of new names was created with whatever information about the author that could be gathered from the bibliographic record (research assistants were trained to record as much information as was found in the original work) and then a manual search for an established name authority using the sources noted above was conducted. Once a name authority was obtained or created it was added to the authority record of the working file along with the authority source code. When the records in the working file were completed, the file was appended to the main name authority file. Working files were created throughout the data gathering stage of the project and authorities were added to the main file as soon as possible to keep it up-to- date. Revisions to the main file were made when mistakes, such as typographical errors, or author names having the same authority, were found.

Apart from the name authority records, the author name authority file includes cross-references ("see" references and "use" references). "See" references refer a user from a variant name to the "legal" authority name and they were obtained from the established authority sources or were created using the variant author names encountered in the original publications (from which bibliographic entries were prepared).

When the tiles of the Bibliography were in various stages of completion (the files *were* not merged until late in the project and were not readily available online for perusal). duplicate name authorities were sometimes inadvertently created for perusal). duplicate name authorities were sometimes inadvertently created. Rather than eliminate a duplicate authority record (requiring extensive searching of the bibliographic tiles to locate and change the authority numbers), one of the duplicate authority records was retained as the "legal" authority and the other was demoted to a "use" reference. When working copies of the author catalogue *were* generated, "use" references provided a means for accounting for an error and pointing to the "legal" name authority.

The link between the author name authority file and the bibliographic records, as stated earlier, is the author number (see Figure 2). This unique number was entered in the author field of the bibliographic record (works with multiple authors were assigned multiple author numbers) and this translated to the author name authority when the author catalogue generation program was run.

Authority control applied to subject descriptors
More and more attention is being focussed on authority control as applied to subject heading and subject classification tiles (see Bastow, 1984; Grosch, 1972; McKinley, 1972; O'Neil, 1982; and Way, 1968). For this Bibliography, subject access was a necessary priority. However, the decision to provide subject access posed some difficulty largely because the Bibliography includes entries for works published over three centuries during which science and technology and their associated terminology underwent dramatic development.

Rather than adopt or undertake major revisions to an existing subject classification scheme, the decision was made to create a scheme designed both to treat the topics of the literature included in the Bibliography and to meet the subject requirements of potential users.

The subject classification scheme, discussed in more detail elsewhere (MacDonald and Richardson, 1985), will be briefly described here. The scheme was built in several stages employing a combination of two basic methodologies, namely, the 'empirical' approach in which the candidate descriptors were extracted from the literature, and the 'committee' approach in which the descriptors were selected by a panel of appropriate subject experts. The subject classification scheme is composed of three parts, an outline of all the descriptors, a thesaurus of terms assigned to each descriptor, and an index of all the descriptors and thesaural terms in the scheme.

The subject classification scheme was designed to serve two purposes, namely, an authority structure for the subject catalogue of the final product, and, a mechanism for consistent subject assignment by a single individual or among research assistants who were scanning the literature and assigning subject authority numbers to bibliographic records. Accordingly, much of the subject classification scheme was built prior to bibliographic transcription work and then was revised as needed based on the literature encountered or an understanding of the needs of users was extended.

The subject classification scheme contains a total of 164 descriptors consisting of major classes and subdivisions arranged into three categories — those dealing with the sciences, those treating technology, and those that overlap both categories, *e.g.,* education, exhibitions, museums, and societies. For each descriptor an authority record was created and contained up to live elements:

a. a subject authority number unique to the descriptor
b. the authorized descriptor
c. a definition and/or scope note
d. a cross-reference ("see also" reference)
e. terms encountered in the literature assigned to that descriptor

Each authority record always contained elements a, b, and e, and selected records contained elements c and d. While the first four elements of the authority record are self explanatory, the fifth may need an explanation. In order to more clearly define the subject descriptors, terms that were found or night be found in the literature were assigned to each descriptor authority record. Having such a list of terms was deemed more helpful than providing definitions alone. Approximately 2,000 terms were assigned to the descriptor authority records and each of the terms were entered in the index of the complete subject classification scheme (see Figure 3).

Each record was entered into the subject descriptor authority file according to the subject authority number assigned to ensure that the record (and its descriptor) fit into the appropriate place in the subject classification scheme (see Figure 4). Subject authority numbers were assigned to each bibliographic record of the database (maximum of six per record) and the numbers formed the link to the subject descriptor authority file (see Figure 2).

As the subject descriptor authority file was built, revisions to records already in the file were made as needed. In most cases, the modifications were of a minor nature; in others, the revisions required reworking of one or several related descriptors.

Authority control applied to serial titles
Serial title authority control was established to meet the following needs:

a. control for consistent form of titles
b. control for title changes
c. control for typographical or transcription errors
d. provision of a list of serials to be searched

A preliminary list was created from titles of serials noted in the pilot project and/or found in the *Union List of Scientific Serials in Canadian Libraries* (Canada Institute for Science and Technology Information, 1979, 1984, 1985). Title variants noted in the *Union List of Scientific Serials* and foreign serial titles were also added to the list and serial numbers were assigned to each title (original or variant). Foreign serial titles were obtained from experience with the pilot project and from information obtained as the work of the current project continued, when it was assumed that these were likely to include a Canadian contribution or topic.

Each entry of a serial authority record consisted of four elements or fields:

a. a serial authority number unique to each title
b. the authorized serial title
c. an annotation
d. a title code indicating a primary title or variant

As the titles in the authority file were searched for bibliographic information, the title as it appeared on the title page of the serial was verified against the title in the authority file. Any errors in the pre-established title authority were then
corrected.

New titles were added to the authority tile if a serial was searched and an authority did not yet exist in the file. The serial title authority was created from an exact transcription (except for capitalization and punctuation) of the title as it appeared on the title page of the serial itself. Sometimes the titles or title variants were not found in the *Union List of Scientific Serials* as not all title changes are documented in that publication.

At the end of the data gathering stage, a link was established among selected serials, using the serial authority numbers and the title code, to show the relationship between original titles of serials and all their subsequent title changes. This "List of Serials Searched," published with the Bibliography, shows each serial title and all its variants as well as providing a "see" reference from a variant title to the original serial title. The *Union List of Scientific Serials* has yet to do this consistently.

The serial authority number served as the link between the bibliographic records and the serial title authority file (see Figure 2). The serial authority number entered in the serial field on the bibliographic record matched the number in the serial title authority file, and when the catalogue generation programs were run, the appropriate serial title was printed in the bibliographic entries.

Discussion and conclusions
Authority control employed in the construction of this Bibliography was a necessary component of the overall design. While it is recognized that in a database available for online searching it is not absolutely necessary to have standardized forms of names, subject descriptors, or serial titles, provided that links between all possible variants are provided in the searching structure, for the published version of this Bibliography, authority control could not be overlooked. As the number of records in the database grew, the benefits of the authority files became increasingly clear. Time loss due to simple transcription errors (as in the case of the serial title, *Fern bulletin,* incorrectly entered as *Farm bulletin)* was prevented simply because the only necessary correction was a change in the authority file and not to all the records containing the errors which were scattered throughout numerous files (remembering that bibliographic data files Were not merged into one large database until much of the editing had been completed). And further, from a different point of view, authority control was considered a necessity to allow for maximum efficiency in browsing the catalogues of the final product. A user can search entries under an author name and be assured of the final product. A user can search entries under an author name and be assured that all works b_y that author are located there. Similarly, a search under a subject descriptor ensures that all entries in the Bibliography have been collocated under the descriptor.

While authority control was deemed a necessity, its application was not always easily carried out. Authority files are largely transparent in databases or $onlin_e$ catalogues which leads many to underestimate the resources required to generate and maintain such files (see Clark, [1]985-86). For this Bibliography many hʋn_985-86). For this Bibliography many hundreds of hours had to be dedicated to the research and other work required to create the authority files. For the most part, the full implications of this was not considered when the project proposal was prepared. For a retrospective database, such as this Bibliography, probably more time and resources are needed to create authority files simply because much of the work is original. Existing published authority tiles often prove to be quite inadequate.

Because of its size, the author name authority file required the greatest dedication of resources to build and maintain. A number of observations need to metion of resources to build and maintain. A number of observations need to be noted here concerning the literature cited in the Bibliography. Of the total number of records in the database, a large percentage were published anonymously (over 24,000 of the 58,109 records). This may seem an unusually high figure but it reflects the publishing pattern of the Victorian period during which "it has been estimated that ... close to seventy-five percent of the articles ... published in monthlies and quarterlies were anonymous or pseudonymous" (Hiller, 1978, 124).

For the remaining records in the Bibliography, all of which had identifiable authors, Table 1 and 2 and Figure 5 provide a clearer picture of the composition of the author authority file. Of the 10,910 records in the authority file, 824 were "see" references or "use" references. Of the remaining 10,086 records, Table 1 and Figure 5 show the sources of the name authorities. The figure that stands out clearly is the number of name authorities that had to be uniquely established by the staff of the Bibliography. Over fift$_y$ percent of the name authorities are new. Of the established name authority sources, *Canadíana Authorities Canadíana Authorities* and the *National Union Catalogue* proved the most helpful, providing 24.6% and 20-5e of the name authorities, respectively. While the number of authorities obtained from the *National Union Catalogue* is lower than that for *Canadíana Authorities,* it may not mean that the latter is a better source of name authorities since it was given first priority when searching for an authority. Other established name authority sources were less helpful, although it must be remembered that this reflects to a certain extent the fact that these sources were given a lower priority when searching for an established authority. Established name authorities could not always be accepted with 100% confidence. In a few cases textual and bibliographic information alerted the staff of the Bibliography to a possible error in the authority assigned by a name authority agency and on checking biographical sources the error was confirmed. In such cases the established name authority was rejected and a new name authority created.

While more than half of the name authorities were uniquely established for the Bibliography, it is also instructive to note that less than five percent of these names $_c$ould be confirmed in biographical sources (see the list in Appendix A). This low number of authors featured in any biographical source, however, did not eliminate the many hours of searching for such information. The small number also indicates that many of the authors in the database were not important enough to warrant published biographies and points to the low priority workers in science and technology have received in Canadian historical research. The lack of information to aid in the establishment of authorities for so many names was a major contributing factor to the extensive time and resources needed to generate the name authority file.

Other data emanating from the name authority file include the number of corporate authors, women authors, and pseudonymous authors (see Table 2). Establishing name authorities for corporate authors, particularly for manufacturing firms, was sometimes done with difficulty. The history of technology in Canada is still largely unwritten and this means that usually little historical information on manufacturing firms was available to consult- Women authors also posed some difficulty simply because of the convention of signing their names in the form "Mrs. John Doe." The total number of women authors in the authority file is not large—only 170 out of 10,086 or 1.7%. Of these, twenty-three signed their name in the form "Mrs. John Doe" and of these, only six were confirmed in established name authority sources. To establish a name authority for pseudonymous authors almost always required more research than with known author names as it was not always obvious if a pseudonymous author was using his own name or a pen-name. As a case in point, one prominent bee-keeper, William Clarke, used the pseudonym "Lindenbank" or "Linden Bank." Name authorities were created for both forms of the pseudonym until a chance conversation about the role of linden trees in beekeeping alerted staff members of the Bibliography to the possibility that "Linden Bank" was a pseudonym and an attentive research assistant found information that confirmed that William Clarke often wrote under that pseudonym. Fortunately, the number of pseudonyms was not large (only 140 out of 10,086).

One aspect which was considered during the development of the authority files concerned the level of comprehensiveness, i.e., whether sufficient cross-references and notes were provided. "See" references

generated for the author name authority file were transferred to the author catalogue. Cross-references in both the subject descriptor authority file and the serial title authority file are included in the microfiche copies of the published Bibliography. The resolution of the difficulty of determining how many cross-references are sufficient could only he based on information gained while collecting bibliographic data and on "test" users' experience with a working copy of the database. One is never entirely sure that the best resolution of the difficulty has been adopted. The name authority file includes 692 "see" references or a ratio of "see" references to name authority records of 1:14.5. Although some research has examined the importance of cross-reference structure *(e.g.,* Jamieson, Dolan, and Declerck, 1986) much of the literature on name authority files is not helpful in deciding whether this ratio is adequate.

The creation of the subject descriptor authority file, much like the name authority file, required extensive dedication of resources, particularly at the beginning of the project and later at the end of the first phase of data collection when over 24,000 records had been accumulated. Because the subject descriptor authority file was built as a subject classification scheme with a defined structure, revisions to one subject descriptor authority record required careful consideration of other related authority records. Some subject descriptors were subdivisions of others and revisions needed to take into account descriptors at the same level and at higher or lower levels. Revisions of this nature were not purely mechanical.

A number of authors (e.g., Piternick, 1982; and Williams and Lannon, 1981) have commented on the need for standardization of serial titles in databases and have suggested using ISSN or CODEN codes in bibliographic records which would translate to the full title when a user conducted an online search. Most early periodicals, however, do not have an ISSN or CODEN code. And further, at least for this Bibliography, the major difficulty in creating an authority file of serial titles concerned the deficiency of detailed information in published lists of serial titles. Early Canadian periodicals have not been studied to any great degree and, as a result, descriptions of the publishing history of many periodicals is incomplete. Very early in the project it was discovered that titles and title variants encountered in bibliographic transcription work were not accounted for in the standard lists of serial titles. It, therefore, became necessary to develop detailed records on the publishing history of each periodical under review. These records became the basis of the serial title authority file. An authority record was created for each serial title based solely on the title as transcribed from the title page of the serial. Much of the work associated with creating the serial title authority file was unpredicted at the beginning of the project and during the later editing stages of the database it became necessary to revise the design of the serial title authority file so that a published version could be created in which the relationship between a primary title and all its title variants could be clearly shown- The serial title authority file contains 132 "see" references from variant titles to primary titles or a ratio of "see" references to primary titles of 1:1.6, an indication of the high number of changes serial titles underwent.

Robert Burger has written that "unlike bibliographic records, there is no standard code like AACR2 for the formation of the content of authority records"standard code like AACR2 for the formation of the content of authority records" (Burger 1985, 12). Until very recently there was no agreement on the elements or fields that should be included in an authority record and neither was there agreement on rules for the content or style of the data in each field- International standards are only now being agreed upon (and, at any rate, were not available when this project was begun). Relying on established authority sources can mean the use of conflicting styles. And further, because of the lack of generally accepted standards, designers of authority tiles have been left to create their own which invariably leads to confusion and obstacles for use of authority files in different applications. The lack of standards on which to rely posed a continuing difficulty in the construction of the authority files of this Bibliography.

While some authors have suggested that integrated authority files have greater value than separate files (Clark, 1985-86, p. 137), integration is only of value if there are common elements between types of authority files as might occur with name authorities also functioning as subject authorities. Such integration is also only feasible if there are standard components in authority records- In any case, the authority files for this Bibliography were

best maintained as separate files (a relationally DBMS was used) because there was no commonality of content and the authority files were applied to different fields of the bibliographic record not just access heading alone.

Evaluation of authority systems is a very recent development- Burger has written that he has not found "any evidence of comprehensive evaluation of authority data" (Burger, 1985, p. 39). Discussions of authority files often describe activity intended to ensure the correctness or legality of the authority data and its format. This is, of course, a necessary part of authority work and many attempts were made to ensure a high degree of accuracy in the authority files of this Bibliography. In the author name authority file less attention than might be expected was given to precise legality of format simply because of the different standards of the established authority sources from which name authorities were extracted. Adopting name authorities with slightly different standards of format did not lead to problems in generating the catalogues and, therefore, it was not considered necessary to apply a precise legal format to all name authorities.

A number of evaluation questions that could be addressed relate to the growth of the authority file. Bibliometric analysis has been applied to the growth of information. It may be possible to use this methodology in analysis of authority files- In a discussion of several name authority systems, Burger has cited the ratio of authority records to bibliographic records in each system. At the time he wrote (mid-1984), the ratio for OCLC was 1:11; for RLIN I :6; for the Library of Congress 2:5; for the New York Public Library 1:2; for MELVYL 2:1 and for the Washington Library Network 1:1. At the moment, however, the ratio is really meaningless- Information provided b_y the various authority systems makes it very difficult to ensure that the same thing is being measured in each system. For this Bibliography the ratio is about 1:3.4. But it, too, has little meaning. For example, what weight should one give to works with multiple authors? Should each author be credited with full value or some percentage of full value? How does the question of weight relate to understanding the ratio of authority records to bibliographic records? Further study is required to determine the importance and meaning of the ratio of authority records to bibliographic records.

A number of implications arise from this study of authority control for any future work with retrospective databases.

1) Until recently, there were no comprehensive guides to authority work. Difficulties encountered in the design of authority records and files of this Bibliography may be avoided in the construction of other retrospective bibliographics if agreement is reached on standards for the formation of authority records.
2) The use of authority files greatly reduced the number of potential transcription errors in this Bibliography. In addition, by using authority control, it is estimated that the size of individual bibliographic records was reduced by 40e7, resulting in more efficient use of computer memory.
3) Authority control was a necessary component of the design of this Bibliog- raphy, but, the extent of resources needed to create the authority files was unpredicted. Retrospective databases, by the very nature of the information they contain, probably require greater resources for their creation than data- bases of more current literature. Designers of retrospective databases *need* to keep this in mind.
4) No doubt additional retrospective databases will be created to meet the needs of historical research- Designers of such databases should not have to recreate authority tiles. A mechanism needs to be made available to integrate and make available retrospective authority files much like those provided b_y bibliographic utilities.

To bring the paper to a conclusion reference must be made to the intriguing case of Joseph Tweddie and J. Tweddle. Were these gentlemen one and the same individual? -a question that puzzled us for some months. In the last two decades of the nineteenth century Joseph Tweddie of Stoney Creek, Ontario, wrote numerous articles on fruit-growing for the *Canadian Horticulturist.* During the same period the same journal also carried articles by a J. Tweddle of Stoney Creek. Many publications were consulted in an attempt to determine whether Tweddle was an actual individual or a ghost created by a careless typographer. Was it a case, as in *Alice in Wonderland,* of Tweedle-Dum echoing the work *of* Tweedle-Dee? While we decided on the basis of the limited

information available that Tweddle was not Tweddie, we leave it to some prospective user of the Bibliography to unravel the mystery and inform us whether "Tweddle-Dum" was really a pseudonym for "Tweddie-Dee!"

Appendix A: Select list of biographical sources used to assist in establishing author name authorities

BAILEY, T.M., ed., *Dictionary of Hamilton Biography.* Vol. 1. Hamilton, Ontario. Dictionary of Hamilton Biography, 1981.

The Canadian Almanac, and Miscellaneous Directory, for the Year 1895. Toronto: Copp, Clark Company, 1893.

The Canadian Almanac and Miscellaneous Directory for the Year 1900. Toronto: Copp, Clark Company, 1899.

CHARLESWORTH, Hector, ed., *A Cyclopaedia of Canadian Biography.* Toronto: Hunter-Rose Company, 1919.

49 Authority control

Dictionary of Canadian Biography. Vol- 1- 1000 to 1700. Toronto: University of Toronto Press, 1966 and subsequent volumes.

ENROS, Philip C., corp., ENROS, Philip C., corp., *Biobibliography of Publishing Scientists in Ontario Between 1914 and* 1939. Thornhill: HSTC Publications, 1985.

GREENE, B.M- *Who's Who and Why l919-20.* Vancouver: International Press, c. 1912.

Who's Who in Canada (Registered) Including the British Possessions in the Western Hemisphere 1928-29. Toronto: International Press, c. 1924.

KENNEDY, John F., ed., *Who's Who and Why in Canada (and in Neofoundlanad).* vol- 1., Vancouver: Canadian Press Association, 1912.

MORGAN, Henry James, ed., *The Canadian Men and Women of the Time: A Hand- Book of Canadian Biography.* 1st ed. Toronto: William Briggs, 1898.

. *The Canadian Men and Women of the Time: A Hand-Book of Canadian Biography of Living Characters.* 2nd ed- Toronto: William Briggs, 1912.

PARKER, C-W., ed., *Who's Who and Why: A Biographical Dictionary of Notable Men and Women of Western Canada,* vol. 2., Vancouver: Canadian Press Association, 1912.

Who's Who and Why: A Biographical Dictionary of Men and Women of Canada and Newfoundland, vols. 6 and 7, 1915-16. Vancouver: International Press c. 1914.

People of Ontario 1600-1900- 3, vols. London, Ont.: The Genealogical Research Library, 1984.

ROBERTS, Sir Charles G-D-, and TUNNELL, Arthur L., eds., *A Standard Dictionary of Canadian Biography: Canadian Who Was Who 1875-1937.* (Ill.) ed., 2 vols. Toronto: Trans-Canada Press, 1934 and 1938.

STEPHENSON, HSTEPHENSON, H.H., ed., *Who's Who in Science International 1914.* London: *J.* and A. Churchill, [1914.

STORY, Norah. *The Oxford Companion to Canadian History and Literature.* Toronto: Oxford University Press, 1967-

THOMAS, Arnold W., ed., *The Canadian Almanac and Miscellaneous Directory for the Year 1905.* Toronto: Copp, Clark Company, 1904.

WALLACE, W. Stewart, comp., *A Dictionary of North American Authors Deceased Before 1950.* Toronto: Ryerson Press, 1981.

The Macmillan Dictionary of Canadian Biography. 4th ed., rev., enl., upd., by W.A. McKay. Toronto: Macmillan of Canada, 1978.

Table 1. Sources for name authorities

Name authority sources	Number of authorities	% of authorities
The Project (Scitec)	5,147	51.0%
Canadiana Authorities (NLC)	2,484	24.6%
National Union Catalogue (NUC)	2,070	20.5%
Canadiana Monographs 1867-1900 (Retro)	216	2.2%
Canadian Institute for Historical Microreproductions (CIHM)	89	0.9%
British Library Catalogue (BL)	48	0.5%
Library of Congress Authorities (LC)	23	0.2%
University of Western Ontario Library Catalogue (OLU)	9	0.1%
TOTAL	10,086*	100.0%

* This figure excludes cross-references, which totalled 824

Table 2. Contents of name authority file

Item	Number
Cross References	824
Corporate Authors[1]	410
Pseudonymous Authors[1]	140
Women Authors[2]:	170
Women author names in the form, "Mrs. John Doe"	23
Names in the form, "Mrs. John Doe" Available in established authority sources	6
Authority names established by the Bibliography and confirmed in biographical sources	229

[1] Based on a sample of the author name authority file.

[2] Total of all recognizably female names.
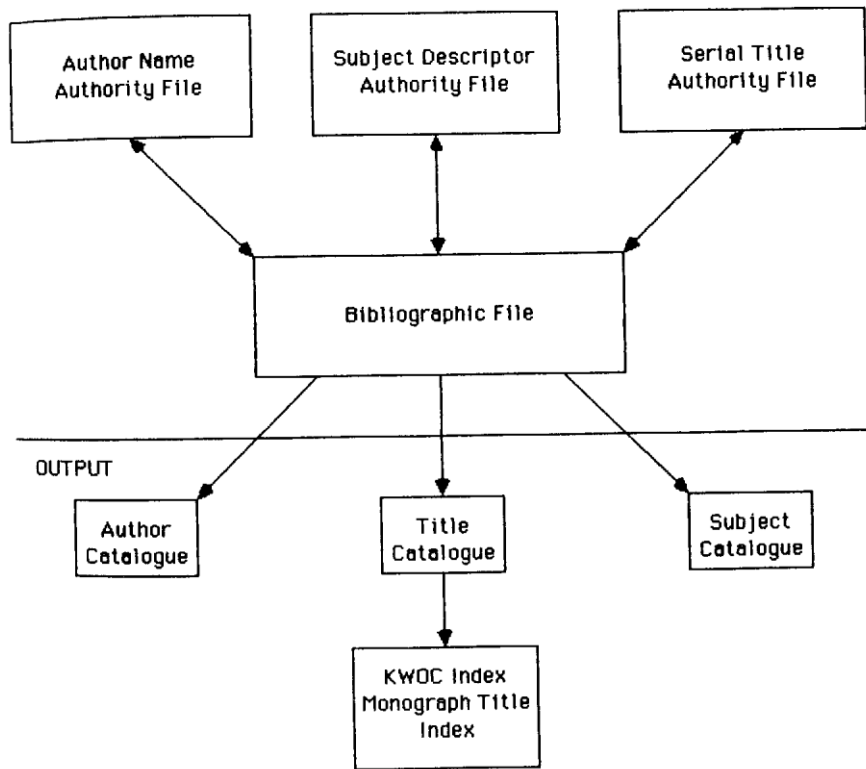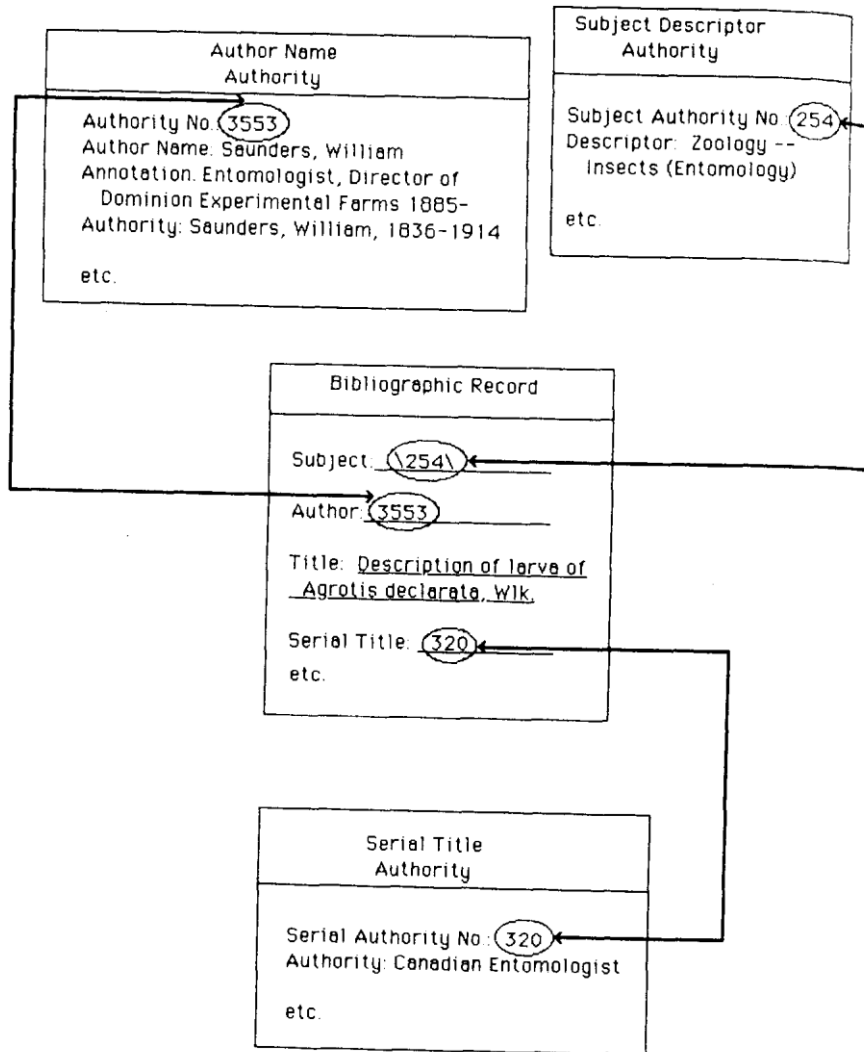
Figures

Figure 1 -- File Structure of the Database

Figure 2 -- Linking Structure of the Database

Figure 3 - Sample of Subject Descriptor Authority Records

Note: In this example, the second authority record is a
        subdivision of the first.

Authority no: 50
Authorized descriptor: Botany
Cross reference: See also related categories under Geology:
                         *134 - Palaeontology and *136 - Palaeobot-
                         any; See also *307 - Agriculture-Economic
                         Botany.
Terms assigned: Botanical Discoveries
                         General Works
                         Geographical Distribution
                         Plants (in general)

Authority no: 51
Authorized descriptor: Lower Plants (Algae, Fungi, Lichens,
                         and Mosses)
Cross reference: See also related categories *175 -
                         Oceanography and *176 - Study of
                         Inland Waters (Including Limnology)
Terms assigned: Bryophyta
                         Cryptogamia
                         Mycology
                         Protophyta
                         Sporulation
                         Sporophyte
                         Spore

Figure 4 – Sample of the Subject Descriptor Authority
File Showing Filing Order

Note: Only the outline of authority records
is shown here

Scientific Topics

50   Botany
51   Lower Plants (Algae, Fungi, Lichens, and Mosses)
53   Forestry
54   Higher Plants (Pteridophytes and Spermatophytes)
56   Materia Medica
58   Plant Classification
59   Physiological and Morphological Botany

Technological Topics

370   Materials and Processes
371   Ceramics
372   Chemical Industry
373   Explosives
374   Photography
376   Fuels
377   Lumber Industry and Woodworking
378   Mills, Factories, and Foundries
380   Mining
381   Underground Operations and Surface Mining
382   Milling and Crushing
383   Smelting and Refining Mills
384   Metallurgy
385   Metalworking
386   Paper Making and Printing
388   Textiles and Clothing

Figure 5 -- Sources of Name Authority By Percentage
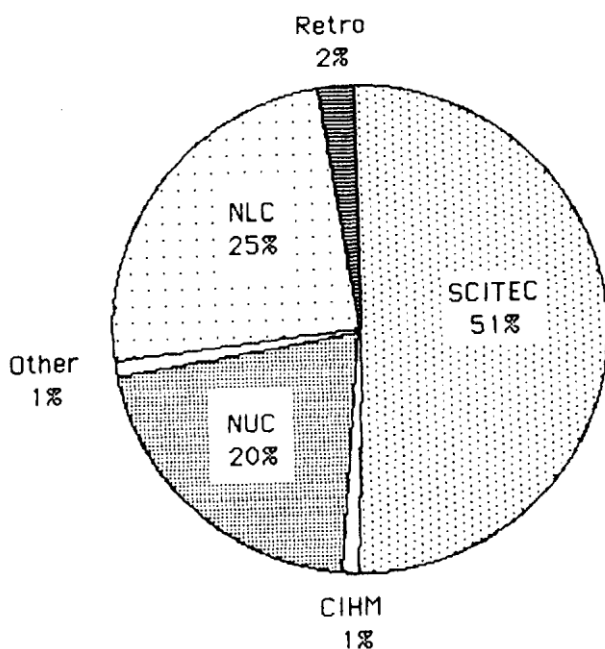
Key:
CIHM — Catalogue of the Canadian Institute for Historical Microreproductions
NLC — Canadiana Authorities of the National Library of Canada
NUC — The National Union Catalog Pre-1956 Imprints
Retro — Canadiana Monographs 1867-1900 of the National Library of Canada
SCITEC — Science and Technology in Canadian History: A Bibliography of Primary Sources to 1914.
Other — Library of Congress Authorities, Library Catalogue of the University of Western Ontario, British Library Catalogue.

References

All D, Larry. "Authority Control: An Eighty-year *Rcview"Library Resource cal Services* 26 (October/December 1982): 319-30.

Bastow, Stan. "ORACLE Digests LC Subject Authorities at State Library of Queensland," *LASIE* 14 (March/April 1984): 22.

BURGER, Robert H. *Authority work: the creation, use, maintenance, and evaluation of authority records-* Littleton, Colorado: Libraries Unlimited, 1985-

CANADA INS ГI 1110 FOR SCIENCE AND TECHNOLOGY INFORMATION. CANADA INS ГI 1110 FOR SCIENCE AND TECHNOLOGY INFORMATION. *Union List of Scientific Serials in Canadian Libraries.* Ottawa: Canada Institute for Science and Technology Information, 1979 and subsequent editions.

CLARK, Doris Hargreth. "On Becoming an Authority on Authorities: A Working Bibliography," *Florida Libraries* 33 (January/February 1983), pp. 13-18.

CLARK, Doris Hargreth. "Authority Control: 'Issues and Answers'," *Technical Services Quarterly 3 (1985-86):* 127-40.

Ci ГI ГI IR, Charles A. *Rules for a Dictionary Catalog.* 4th ed., Washington, D-C.: Governmen*t* Printing Office, 1904.

ENROS, Philip C., corp., *Biobibliography of Publishing Scientists in Ontario Between [1]914 and 1939.* Thornhill: HSTC Publications, 1985.

GROSCH, GROSCH, A.N. "Computer-based Subject Heading Authority Files at the University of Minnesota Libraries," *Journal of Library Automation* 5 (1972), pp. 230-43-

HILL ER, Mary Ruth. "The Identification of Authors: The Great Victorian Enigma." *in*

Victorian Periodicals. A Guide to Research, pp. 1 I23]-148. Ed, J. Don Vann and
Rosemary T- Van Arsdel. New York: The Modern Language Association of America, 1978.

JAMIESON, Alexis .1., DOLAN, Elizabeth, and DECLERCK, Luc. "Keyword Searching vs.
Authority Control in an Online Catalog", *Journal of Academic Librarianship* 12 (1986), pp. 277-83.

JARRLL, Richard, and RODS, Arnold. A *Bibliography for Courses in the History ofCana-
dían Science, Technology and Medicine.* 2nd rev. ed., Thornhill: HSTC Publications, 1984-

MACDONALD, Bertrum 11., and RICHARDSON, R. Alan. *Preliminary Bíbliographical Inven-Preliminary
Bibliographical Inven-
tory 0/ Sources in the History of Science, Technology, and Medicine in Canada to the Tiren-
tieth Century.* London: Department of History of Medicine and Science, University
of Western Ontario, 1981.

. "Science and Technology in Canadian History: A Report on the Construction of a Retrospective
Bibliography." Paper presented to the Association for the Bibliography of History, San Francisco, December
1983.

• "The Development of Subject Classification Scheme for the • "The Development of Subject Classification
Scheme for the *Bibliography of Canadian Science and Technology to 1914-" Papers of the Bibliographic
Society of Canada 24 (1985),* pp. [86]-103.

MALINCONICO, MALINCONICO, S. Michael. "Bibliographic Data Base Organization and Authority File
Control." In *Authority Control: The Key to Tomorrow's Catalog. Proceedings of the 1979 Library and
Information Technology Association institutes,* pp. 1-15. Edited by Mary W. Ghikas. Phoenix, Arizona: Oryx
Press, 1982.

MCKINLEY, J. "Concerning Subject Authority Catalogues," *Library Resources & Technical Services* 16 (Fall
1972), pp. 460-65.

O'NEILL, Edward T. "Computer Generation of a Subject Authority File," O'NEILL, Edward T. "Computer
Generation of a Subject Authority File," *Information Interaction: Proceedings of the 45th ASIS Annual
Meeting, 1982, Columbus, Ohio* 19 (17-21 October 1982), pp. 220-23.

PITERNICK, Anne B. "Standardization of Journal Titles in Data Bases," *Journal of the American Society for
Information Science* 33 (1982), p. 105.

RICHARDSON, R. Alan, and MACDONALD, Bertrum H- Bertrum H- *Science and Technology in Canadian
History: A Bibliography of Primary Sources to 1914.* Thornhill, Ontario: HSTC Publications, 1987.

ROCHEFORT, Alain. "Le contrôle Alain. "Le contrôle Alain. "Le contrôle d'autorité," *Documentation et
bibliothèques* 31 (octobre-décembre 1985), pp. 151-57.

STEVLNSON, H.H., STEVLNSON, H.H., ed., *Who's Who in Science International 1914.* London: J. and A.
Churchill, [1914]-

WAY, William. "Subject Heading Authority List, Computer Prepared," American *Documentation* 19 (1968),
pp. 188-99.

WILLIAMS, M-E-, and LANNON, L. "Lack of Standardization of the Journal Title Data Element in
Databases," *Journal of the American Society for Information Science* 32 (1981), pp. 229-33.