



ARCHIVED - Archiving Content

Archived Content

Information identified as archived is provided for reference, research or recordkeeping purposes. It is not subject to the Government of Canada Web Standards and has not been altered or updated since it was archived. Please contact us to request a format other than those available.

ARCHIVÉE - Contenu archivé

Contenu archivé

L'information dont il est indiqué qu'elle est archivée est fournie à des fins de référence, de recherche ou de tenue de documents. Elle n'est pas assujettie aux normes Web du gouvernement du Canada et elle n'a pas été modifiée ou mise à jour depuis son archivage. Pour obtenir cette information dans un autre format, veuillez communiquer avec nous.

This document is archival in nature and is intended for those who wish to consult archival documents made available from the collection of Public Safety Canada.

Some of these documents are available in only one official language. Translation, to be provided by Public Safety Canada, is available upon request.

Le présent document a une valeur archivistique et fait partie des documents d'archives rendus disponibles par Sécurité publique Canada à ceux qui souhaitent consulter ces documents issus de sa collection.

Certains de ces documents ne sont disponibles que dans une langue officielle. Sécurité publique Canada fournira une traduction sur demande.

*/Indexing and abstracting standa
a CRIM/DOC reference information*

Creating A Criminology Documentation
Data Base (CRIMDOC Data Base):
Phase III: Final Report

CENTRE OF CRIMINOLOGY



UNIVERSITY OF TORONTO

Z
699.5
.C76
M3
1982
pt.3

Z
699.5
.C76
M3
1982
pt.3

*/Indexing and abstracting standards for
a CRIM/DOC reference information bank/*

Creating A Criminology Documentation
Data Base (CRIMDOC Data Base):
Phase III: Final Report

prepared by

Catherine J. Matthews
Centre of Criminology
University of Toronto

under contract to

Solicitor General Canada
Contract No. 7270-29



August 1982

Copyright of this document does not belong to the Crown.
Proper authorization must be obtained from the author for
any intended use.
Les droits d'auteur du présent document n'appartiennent
pas à l'Etat. Toute utilisation du contenu du présent
document doit être approuvée préalablement par l'auteur.

TABLE OF CONTENTS

Abstract	<i>ii</i>
Introduction	i
General Background Assumptions Concerning CRIMDOC	3
Summary of Report #1: Indexing and Abstracting Standards	6
Summary of Report #2: Indexing Vocabulary	9
Cost Factors in Creating the Data Base	
(i) Creating the Vocabulary	14
(ii) Creating the Pilot Data Base	15
(iii) Creating CRIMDOC in a Self-Support Environment	15
(iv) Creating CRIMDOC in a Network Environment	16
Networking	17
Document Delivery	20
Summary and Recommendations	23
Appendix A: Persons Contacted	29
Appendix B: Footnotes	31
Appendix C: Bibliography	33

ABSTRACT

Creating a Criminology Documentation Data Base (CRIMDOC Data Base):
Phase III: Final Report. Prepared by Catherine J. Matthews,
Centre of Criminology, University of Toronto, under contract
to Solicitor General Canada, Contract No. 7270-29. August, 1982.

This is the third and final report to be prepared under contract to the Solicitor General Canada on considerations involved in the establishment of a bibliographic data base of Canadian criminological research and documents. The two earlier reports covered indexing and abstracting standards for CRIMDOC and indexing vocabulary for information storage and retrieval, and are summarized in this third report. This third report particularly emphasizes the administrative considerations involved in the creation and operation of a data base.

While various existing methods of information retrieval, including on-line use of NCJRS (National Criminal Justice Reference Service) are presently used by the Canadian criminal justice research community, they do not meet all the needs, and consequently there is a perceived need for a uniquely Canadian resource file. The proposed CRIMDOC data base (Criminology Documentation) would likely contain bibliographic information from selected Canadian journals, unpublished research reports and research in progress, theses, chapters in edited texts, conference papers, government reports and selected monographs.

The priorities now before the Ministry of the Solicitor General include the need to issue an internal report clarifying objectives and prepare a time schedule for the creation of the data base, as well as beginning to budget for the data base. The first and perhaps greatest cost factor in establishing the CRIMDOC data base will be in establishing the indexing vocabulary. After the vocabulary is ready, a pilot data base must be built up, tested and evaluated.

There are considerable advantages for the Solicitor General in creating CRIMDOC in a network environment involving participation by a number of criminology departments and centres, the most significant of which is improved scope of coverage. Network contributions to the data base would also mean a sharing of costs, and less time-lag in data entry.

Certain significant questions will have to be resolved before a network can materialize, particularly with regard to funding for acquiring the computer hardware necessary to participate. Agreement on the important matter of document delivery must be reached, so that requestors will be able to access full copies of materials cited in the data base.

The CRIMDOC data base is most likely to be successful as an independent file, operated in a network arrangement through a commercial service bureau, with matters relating to project administration, training, thesaurus maintenance and document delivery centralized under the authority of the Library and Reference Centre of the Solicitor General of Canada.

CREATING A CRIMINOLOGY DOCUMENTATION DATA BASE
(CRIMDOC DATA BASE): PHASE III: FINAL REPORT

Introduction

This is the third and final report to the Solicitor General of Canada on considerations involved in the establishment of a bibliographic data base of Canadian criminological research and documents. Throughout this report the data base is referred to as CRIMDOC. This report will serve as a general background document to persons interested in the basic issues involved in establishing such a data base, and its focus will be primarily on administrative rather than technical concerns. Brief, non-technical summaries of the two earlier technical reports¹ are also provided in this report to describe some of the considerations, decisions, and cooperation required to bring CRIMDOC into existence.

Most researchers are presently familiar with several time-honoured methods of information gathering, namely "traditional" library searching of card catalogues and indexes, regular scanning of key academic journals, and "word of mouth" or the "invisible college network" in which researchers keep each other informed of developments in their field. In the last several years a new mode of information retrieval has been made possible by the use of computers. Information which was computer stored and typeset for the production of print indexes was made available for direct on-line searching by computer, and on-line reference searching was born. This automated information retrieval has come of age very quickly, with available data base files proliferating. Advances in computer software have made these information retrieval systems more "user friendly" and less complex

to use so that some can be accessed by untrained searchers. Generally, however, experienced librarians or searchers are better able to select from and use the hundreds of data bases available.

For the criminal justice researcher in Canada there are several relevant data bases which contain information on books, theses, and articles on criminal justice topics. The most relevant data base is NCJRS (National Criminal Justice Reference Service) available through the vendor Dialog. Its Canadian content is limited, and it does not include unpublished research or research in progress, but it is still of great value to researchers. Other commercially available data bases with relevance to the Canadian criminal justice research community include Legal Resource Index, Criminal Justice Periodicals Index, Sociological Abstracts, Psychological Abstracts and Canadian Plains Research, most of which have limited Canadian content. The Register (of Research and Researchers in the Social Sciences) may become available on-line in 1983. Such a variety of files to search, with each having different scope and coverage, varying search strategies, and no one data base adequately meeting the needs of the Canadian criminal justice community, all lead to the desire for a uniquely Canadian resource file which could be accessed on-line. This report will develop some of the assumptions concerning a proposed Canadian criminal justice data base and the advantages it would offer researchers.

A coincidental function of this report is to offer certain observations of an impressionistic nature based on a fairly wide but by no means exhaustive acquaintance with the library and information science literature and libraries and individuals in data base use, creation and administration.

General Background Assumptions Concerning CRIMDOC

In March 1981, a Price Waterhouse Associates study of the Ministry of the Solicitor General's Library² (then called the CRIMDOC Centre, and now known as the Ministry Library and Reference Centre) made Recommendation 4 as follows:

To meet the stated requirements of the CRIMDOC Centre users in the Ministry and those of other users in the Canadian criminal justice community, the CRIMDOC Centre should set up a computerized index to current Canadian programs, projects, research activities, theses, important working papers, criminologists and other significant criminological resources external to the CRIMDOC Centre collection.

Correspondence and communication with Ministry of the Solicitor General Secretariat representatives have focussed the scope of the CRIMDOC data base somewhat more clearly, although not definitively. The proposed data base would likely provide:

- a) regular coverage of specified Canadian journals, regardless of their "retrievability" through other published indexes; as examples, the following titles would be considered: Canadian Journal of Criminology, Canadian Journal of Sociology, Crime et/and Justice, Criminal Law Quarterly, Canadian Criminology Forum, Liaison, etc.;
- b) unpublished research papers, particularly those prepared under contract to the Ministry; this would not include "classified" or restricted reports unless computer software had security levels which limited access to those records;
- c) theses in criminology, as well as relevant theses at the Masters and Doctoral levels in law, sociology, psychology, etc.;
- d) chapters in edited texts or collected works, conference papers and proceedings, government reports and selected monographs. More specific clarification of the scope of this category will have to be articulated, particularly with regard to duplication of access to documents retrievable through the Main Library Catalogue. In this category,

however, books are not a priority, as initially emphasis will be on the more "fugitive" information not accessible elsewhere;

- e) programs, projects and research in progress. Although this category of non-bibliographic items was listed first in the Price Waterhouse Report, the present interpretation is that this data will not necessarily have first priority, but that it will be added to the data base once it is operational. Ideally, however, the plan is to have this data retrievable from the same file as bibliographic information;
- f) the persons inventory of criminologists identified by Price Waterhouse as a desired element in the CRIMDOC data base is best omitted from consideration at this point. At the University of Western Ontario, under the direction of Dr. E.H. Hanis, "The Canadian Register of Research and Researchers in the Social Sciences" is now in the second year of its funding from SSHRC, and at the present time duplicates the concept of a persons inventory proposed for the CRIMDOC data base;
- g) the CRIMDOC file would contain bibliographic references to documents, with abstracts, not the full text of the document itself.

An expectation of the CRIMDOC file is that it would be a "user friendly" system, capable of use by academics, researchers and administrators, without relying on the use of an information intermediary such as a librarian. Initially, however, it may be necessary for the purpose of a pilot project or an evaluation phase to restrict hands-on experience to a limited number of participants. A long term goal of the Ministry in the creation of a CRIMDOC data base is to create a "network" in Canada, with various bodies other than their Ministry Library (e.g., Centre of Criminology Library, University of Toronto; Centre internationale de criminologie comparée; Department of Criminology, Simon Fraser University; etc.) contributing information to the file and retrieving information from it. Ideally, this would result in better national and regional coverage than would otherwise

be possible.

Another possibility in a network environment is that other libraries could participate in "document delivery", that is, getting the actual article, report or document into the hands of the requestor, not just the reference to that document. There are several options in this crucial matter of document delivery, and they are discussed in greater detail later in this report.

A further assumption concerning the CRIMDOC data base is that at some point in the future there will be sufficient interest in the CRIMDOC data base (by every group from correctional workers, police, etc. to public, academic and government libraries), and the data base will have been built up sufficiently that access could be expanded beyond the small number of participants involved in the actual network. The CRIMDOC data base could "go commercial", and be handled by a vendor or service bureau which would make the data base available to any individual on a fee-for-service basis, usually calculated on cost-per-minute of connect time to the processing computer, plus printing costs. This is the manner in which the NCJRS file can be accessed via Dialog (vendor).

Standards for the data base are particularly important because of the possibility of networking and commercial distribution. Internally the file records must be consistent, so that any re-programming required to go commercial would hopefully be minimal. The first and second CRIMDOC reports were somewhat technical, but in the following pages are summarized to distill the administrative concerns of providing standards.

Summary of Report #1: Indexing and Abstracting Standards

A literature search revealed that the concept of bibliographic or records standards actually refers to two concepts where a computer file is concerned. Bibliographic record content standards concern how categories such as author, title, source, imprint, etc. are identified and formatted so that the same information is repeated for each document, and so that print-outs are easily understood by end users. Machine readable format standards are particularly important for communicating with other computers and would be essential in any network or commercial data base environment.

If various persons are to be involved in inputting records into the data base, either in the Solicitor General's Library or in remote network locations, there are several options for record content and data entry. These include

- a) the use of guidelines or standards for data entry;
- b) centralized editing of records; or
- c) data entry on the basis of on-line machine "prompts" with no set standards or editing.

The latter method is the fastest method, but it is not as effective nor in the best interests of this data base, because of the possibility of information scatter.

The second option, centralized editing of records, might be desirable, but it increases time-lag and decreases currency of information, and is costly to execute. The most cost-efficient method for maintaining standards for data entry is to provide clear guidelines to be followed by the person creating the record. Using a format based on the National Library of Canada mini-MARC format, a simplified record content guideline which would accomodate non-bibliographic and well as bibliographic information was

recommended. It is basic enough to be used as input guidelines by non-trained staff, could accommodate, articles, unpublished papers, research and projects in progress, and could be upgraded to a more complex format if changes were required to make the data base commercially available. Formats such as that of the Reference Manual for Machine Readable Bibliographic Descriptions (i.e., the UNISIST Manual) were considered useful but more complex than was ultimately required, particularly in a network environment.

It was recommended that systems design advice be sought on the final design of the format, particularly for "control" fields. To minimize scatter of related works in the data base it would be advantageous if a name authority file could be maintained (either as part of the data base itself, or as a separate file).

Although research is underway³ to standardize inter-computer communications protocols such that all commands would be reconciled to a common format or language, at the present time standards for both record content and communications format are essential elements for a CRIMDOC data base.

It was further advised that concise and informative abstracts should be included in the data base. Although abstracts are expensive to produce, enter and store, they would be of great value to data base users both for providing information additional to the title of a document and for giving greater likelihood of effective retrieval in a data base which allowed free-text natural language searching of complete abstracts. Following critical evaluation of various abstracting standards it was recommended that the American National Standard for Writing Abstracts (ANSI-Z39.14-1979) be adopted for abstracts in the CRIMDOC file. It is clearly in the interest

of minimizing the cost and time-lag of data entry, and maximizing the reliability of abstracts that authors write their own abstracts at source. For those particular documents, data entry staff would only have an editorial capacity to execute, and the data base could be updated much faster.

In consideration of a long term objective of a bilingual data base then staff time, costs and time lag would have to be considered for translation of abstracts.

The following are the actual recommendations of the first report:

General

1. That in creating a data base of Canadian criminal justice information due consideration be given to the overall objective of bringing researchers and information together, and in respect thereof, that consideration be given to the allocation of adequate staff and budget for the various stages, including selection of documents, creation of records, editing of records, processing of documents, information retrieval and document delivery;
2. That a comprehensive collection development policy be prepared for the data base, based on the categories of documents discussed earlier and on which this report is based, namely regularly selected periodical titles, unpublished research papers, Canadian theses, conference papers and proceedings, government reports, chapters in edited texts and collected works, and programs, projects and research in progress;

Bibliographic Standards

3. That clearly defined local standards (such as those appended to the first Report as Appendix D) be utilized for the creation of records, such that a machine readable record format approximating the mini-MARC characteristics be created;
4. That computer specialist design advice be sought on development of format, so that matters of software impact on MARC formats and network protocols to support the transfer of data between computer systems be given consideration;
5. That a separate file be created in machine readable form for the searching of name authorities, or if the software design permits it, that a reference structure for name authorities be built into the main CRIMDOC file;

Abstracts

6. That the proposed CRIMDOC data base contain abstracts which conform to the American National Standard for Writing Abstracts (ANSI Z39.14-1979);
7. That a software design be selected which permits full text searching of a variable length abstract field;
8. That the Solicitor General encourage the creation of abstracts at source in documents prepared under contract to the Ministry by making a bibliographic and abstract style guide an appendix to all contracts signed;
9. That Ministry legal advisors investigate the authorizations required to store abstracts or parts of copyright documents in an information retrieval system;
10. That initially the data base should contain abstracts in the same language as the source document (i.e., English language document, English language abstract; French language document, French language abstract; Bilingual English-French document, Bilingual English-French abstract) but that a long term plan for the provision of a completely bilingual abstract field be considered.

Summary of Report #2: Indexing Vocabulary

This report on indexing vocabulary considered the means of subject access to the data base. Traditional retrieval from library catalogues and indexes has been by means of "controlled vocabulary" - subject headings selected from a list and assigned to documents. Usually these headings have been pre-coordinated (i.e., various words put together to make a "heading" - eg. "juvenile delinquency"), in inverted style (bringing like words and concepts together, as in "Police, Private" instead of "Private police") and their specificity was established at the time of cataloguing (i.e., "pre-coordinated") with subdivision of terms (eg. "Crime and criminals - Research"). Computer searching, however, opens up the opportunity of retrieving information from any of the searchable fields of a

record, using the natural language of the document itself. Searchable fields could include title and abstract in addition to subject heading descriptors. Words could be co-ordinated or put together at the time of the search. This full-text natural language searching opens up great potential for retrieval, as searches can be very specific and have a greater number of access points. The principal disadvantage of using only natural language retrieval is in the scattering of related work having synonymous or near-synonymous vocabulary. For example, in a controlled vocabulary the descriptor "Private police" could be assigned at the data entry stage to control the scatter of other documents using such phrases as "private security", "security guards", "security personnel", "guards", etc.

Any consideration of vocabulary in a data base must be viewed from two perspectives, namely the input/storage considerations and the output/retrieval considerations. The principal advantage of natural language is in saving indexing time and cost, by allowing the computer to store and index every word. Consideration was given to the advantages of cost-saving and speed at the data entry stage, compared to the effectiveness or precision of retrieval using a controlled vocabulary. It was recommended that a controlled vocabulary be used to assign and store subject descriptors in the data base, but that appropriate software be selected so that free-text natural language also be available. In this manner alternative strategies would exist for the searcher.

Once it was established that vocabulary control would be required it was necessary to select a vehicle for vocabulary control. Existing "vehicles" of vocabulary control are the subject heading list, the glossary,

the thesaurus; each has a unique structure and function. Consideration was given to the objectives of the system, the system environment, resources for the creation and management, and the means of evaluation.

Existing subject headings lists and thesauri were evaluated in relation to the above criteria. Subject headings lists presently used in the Ministry Library were rejected as not being "user friendly" enough. They were considered too contrived, not specific enough, and lacking in the provision of generic or hierarchical relationships. Several thesauri, which provided hierarchical relationships were found to be too narrow in scope. The NCJ Thesaurus (National Criminal Justice Reference Service:1980) was considered to be more in keeping with the objectives of the CRIMDOC system, being well designed for "user-friendly" use by persons involved in research, teaching, policy or administrative functions and not assisted by information intermediaries, and permitting highly specific storage and retrieval.

The availability of resources should be considered in the choice of indexing vocabulary. The availability of any suitable existing thesaurus is of great benefit in the decision to use a thesaurus. Staff and financial resources required to build one from scratch are prohibitive and are best avoided. Despite the existence of the NCJ Thesaurus, and its suitability as a "base" vocabulary, considerable staff time would have to be allocated by the Solicitor General for verifying each descriptor (in terms of its relevance in a Canadian context) and then a committee of subject specialists, a translator and/or a lexicographer would have to work on the translation of the vocabulary into French. The manner in which a bilingual thesaurus is best constructed is discussed in the UNESCO/

UNISIST Guidelines for the Establishment and Development of Multilingual Thesauri. It is essential to have a reference standard when the matters of degrees of equivalence of terms, hierarchical and other relationships of terms, etc. are to be resolved.

The single greatest advantage to the Solicitor General would be the opportunity to purchase the magnetic tapes of the NCJ Thesaurus, and although these tapes have never been sold before, representatives of the National Criminal Justice Reference Service (and Aspen Systems Corporation) appear interested in pursuing discussions with representatives of the Solicitor General. Another advantage to the Solicitor General is the opportunity to cooperate with the Translation Bureau, Terminology Directorate of Secretary of State, which has done some background work on vocabulary in the criminal science and parole subject areas.

The adoption of a thesaurus as a means of vocabulary control is an affirmation of a commitment to a special quality of information storage and retrieval. There are very definite advantages to the Solicitor General, such as the availability of a good existing thesaurus, which make the undertaking less costly and less difficult than having to build a thesaurus ab initio. Nevertheless, the costs of buying and translating do exist, and are factors to be reckoned with. In the absence of any specific cost estimates, however, but based on a survey of the literature, observation and use of manual (i.e., print copy) and automated information retrieval systems, and discussions with creators, vendors and users of information storage and retrieval vocabularies, the second report therefore recommended:

General

1. That a software program be used for the CRIMDOC data base which allows for the creation of an inverted file of natural language derived from bibliographic citations and document abstracts;
2. That a controlled vocabulary be developed to store and aid in the retrieval of bibliographic data; and
3. That retrieval from the data base be by means of either the controlled vocabulary, free text searching or a hybrid of both modes.

Specific

4. That a thesaurus be used as the method of vocabulary control, based on the American National Standard Guidelines for Thesaurus Structure, Construction and Use (ANSI Z39.19-1980) (for establishing the integrity of the base thesaurus) and the UNESCO/UNISIST Guidelines for the Establishment and Development of Multilingual Thesauri (for the bilingual thesaurus);
5. That international graphic symbols, such as those found in Lexicographical Symbols Particularly for Use in Classified Defining Vocabularies (ISO-1951-1973E), be considered for use in the thesaurus rather than unilingual alphabetical abbreviations;
6. That the NCJ Thesaurus presently used for storage in and retrieval from the NCJRS data base be acquired in machine readable form to be used as the basis of a Canadian criminal justice thesaurus;
7. That a linguist and/or a lexicographer and a committee of subject specialists revise the NCJ Thesaurus, dropping terms which because of legal definition or other reasons will not be relevant in a CRIMDOC data base, and adding terms to reflect the scope and content of a Canadian criminal justice file;
8. That a small pilot project be undertaken to test the realibility, relevance and retrieval capabilities of the English language vocabulary before the translation of the vocabulary is undertaken, and upon satisfactory results of that project;
9. That a linguist, assisted by a committee of subject specialists, translate the postable descriptors of the English language thesaurus for the purpose of creating a bilingual criminal justice thesaurus.

Cost Factors in Creating the Data Base

Various factors will affect the ultimate cost of establishing a CRIMDOC data base. In the following pages the four principal areas for further cost investigation are outlined and many of the most significant cost factors are discussed.

(i) Creating the Vocabulary

The cost of creating the indexing vocabulary will be a major expenditure for the Solicitor General. Although the National Criminal Justice Reference Service and Aspen Systems Corporation have not as yet provided even a ball-park quotation on the cost of buying the NCJ Thesaurus in magnetic tape format, it is expected that it will be priced to reflect the several years of work required to create and maintain the vocabulary to date. This cost, however, will be much less than the cost of creating a new vocabulary from scratch. The compatible computer software for the thesaurus could be an additional cost in the \$10,000 to \$50,000 range if a program is to be purchased outright. The use of leased software or a computer service bureau would dramatically cut the capital outlay, paying instead only for start-up and consultation, software modifications and storage, with a small on-going charge for services. A distributed system, in which CRIMDOC would share another facility's computer is a more cost-effective method than purchasing an in-house computer.

After purchase of the NCJ Thesaurus tapes considerable staff time would still be required by the Solicitor General to have a committee of subject specialists, a lexicographer and library staff "edit" the NCJ vocabulary to make it a suitable Canadian "base" vocabulary ready for

translation. This committee would then have to work with a translator/linguist to create the French language vocabulary. This editing and translating could be accomplished in one year if the committee met once a month throughout the year, and the lexicographer and linguist also worked full time between committee meetings.

(ii) Creating the Pilot Data Base

A criminologist or librarian could devote approximately one year to building up and evaluating a pilot project data base. Time would be required to select and acquire approximately 500 documents, index them according to format guidelines and assign descriptors based on the vocabulary in the thesaurus. Bibliographic data would have to be coded and entered. Considerable time would be required to do original abstracts for all documents. Furthermore, a means of evaluating the storage and retrieval capabilities should be devised.

(iii) Creating CRIMDOC in a Self-Support Environment

Should the Solicitor General decide to be the sole contributor to the CRIMDOC data base and have sole access to it, they will have to bear the costs completely, as opposed to other arrangements in a network environment. The greatest cost to the Solicitor General would be in trying to establish the CRIMDOC data base as an "integrated" function of the Library, although ultimately this could be of great benefit. As an "integrated" function the acquisition and loan of documents would be recorded in the same system as the available documents, so that the matter of document delivery could be more easily expedited. However, establishing an integrated environment

with automated acquisitions, circulation and reference service is complex, and could take several years to accomplish, and criminal justice researchers may not want to wait that long. An independent file, which could be searched as any other commercial, independent file now used, would be quite acceptable, would cost less, and would take less time to implement.

In a sole-participant or self-support environment all the duties of file creation and maintenance would have to be borne by the Solicitor General's Library staff. Considerable on-going staff-time of professional and clerical levels would be required for document selection and acquisition, indexing, abstracting and data entry into the computer (via word processor, etc.). Vertical files or some such system would have to be maintained for document storage. Staff would be required to conduct searches of the data base, interpret printouts, locate documents, provide loans (both in-house and interlibrary loans) and do photocopying. Professional time will be required on an on-going basis for administration of the service, in terms of handling correspondence with the computer service bureau (or in-house data processing staff, should that alternative be selected), making arrangements for installation and service of hardware, authorizing charge-backs and payments, maintaining the thesaurus, and for updating manuals and guidelines for format and training, or policies regarding services.

(iv) Creating CRIMDOC in a Network Environment

In a network environment involving other libraries and research centres there is the advantage of sharing the work and the costs involved in maintaining a data base. There are, at least initially, certain costs not

involved in a self-support environment. Since many of the possible network participants might not have word processors or terminals to be used for storing information or retrieving it from the data base, and might not have the financial resources to acquire them, some funding would have to be forthcoming to so equip network participants. There will have to be some meetings with network participants to gain their participation agreement, to resolve each member's responsibilities in terms of document selection and acquisition. Funds might have to be made available to support selection and acquisition of documents, as well as to cover photocopying and mailing costs if document delivery is to be provided by network locations. Document delivery is discussed in greater detail in the following pages. Network participants will have to provide staff for acquisitions, indexing and abstracting of documents, hence saving a considerable amount of Solicitor General's staff time. The Solicitor General's Library would be an appropriate body to coordinate and provide training sessions, data base manuals and thesaurus copies for network participants. Such coordination and training is essential if good standards are to be maintained. Further factors concerning the pros and cons of networking are discussed in the following section.

Networking

It is in the best interests of the Canadian criminal justice research community that a network approach be taken for the creation of a CRIMDOC data base. Regional participants would have much better access to research projects and unpublished papers within their own geographic areas than one

person located in Ottawa, and hence national coverage could be greatly enhanced. The sharing of the cost and time for document identification, acquisition, indexing, abstracting and data entry would likely equalize the benefits to be derived from researcher access to the data base. Sharing of the above-mentioned costs and duties would also mean a much more current and up-to-date data base, rather than waiting for data entry from one central location with a limited number of staff.

Certainly there is more complex groundwork involved in the establishment of a network. As was previously mentioned, it is likely that grants to network participants would be essential to provide for computer hardware. Policies concerning responsibility for selection and acquisition of documents will have to be resolved. Agreements, based on document delivery arrangements, will have to be reached with participating members regarding discard and retention periods for documents. As well, editorial discretion for editing records in the data base is best vested in very few hands, and perhaps even ought to be centralized for deletions from the file.

In a network environment the Solicitor General would have to assume responsibility for the training of all persons involved in indexing, abstracting and data entry so that standards are uniform across the data base. This would involve issuing and maintaining a manual for the data base. Ideally a network arrangement would begin with six or fewer participants so that standards are more easily maintained.

Direct maintenance of the thesaurus by network participants is best avoided, and thesaurus maintenance retained as a responsibility of the Solicitor General's Library. Network participants would be encouraged to submit feedback on the vocabulary, although decisions would be centralized. NCJRS maintains the NCJ Thesaurus through the method of indexer feedback

to one lexicographer.

The best locations for network participants would be in Departments or Centres of Criminology corresponding to the major financial support granted by the Solicitor General. In ideal cases these departments have their own criminology libraries which may be flexible enough to accommodate the conditions of network participation. Some departments may not have their own libraries and may rely on their University's main social science and/or law libraries to meet their needs. Some of these large library units may not, however, be in a position to devote suitably qualified staff, time or financial resources to the development of a CRIMDOC data base. In such cases it might be preferable for the Solicitor General to negotiate an agreement directly with a criminology department, bypassing the University library system. A faculty member or researcher could be responsible for contributions to the data base, assisted by secretarial or support staff for data entry. Employment stability and commitment to the data base are essential, as frequent retraining of new participants is costly and could result in poor quality input into the data base.

In a network environment there is considerable advantage in the contact of the contributor (i.e., faculty or researcher actually responsible for that Department's commitment) with colleagues, both for the purpose of collecting references for the file and in alerting them to the advantages of automated retrieval from the file. While certain up-dating of the CRIMDOC file could be done by machine generation of printouts to go to individual researchers for updating, the personal follow-up by peers is extremely valuable.

Document Delivery

The most important consideration in the establishment of the data base is that of "document delivery". It is of little use to the researcher to get a reference and abstract for a document and not be able to acquire a copy of the complete document. There are three specific aspects of document delivery which are closely interrelated: policy, cost and efficiency.

There must be a policy decision made either by the Solicitor General or jointly with network participants concerning who will collect and retain certain documents, journals, etc. Documents acquired in support of the data base must be housed in such a way as they can be directly accessed from CRIMDOC file information. Decisions regarding authorization to loan or photocopy documents must be made for certain categories of materials in particular. There are matters of copyright, funding agency restrictions on research reports, or author's restrictions on photocopying, all of which could become complex if clear policies are not established. For example, in the case of unpublished research reports, will any network participant provide a copy, will the Solicitor General's Library provide copies, or will the inquirer be referred directly to authors or to funding agencies?

If network participants are to provide copies (i.e., either loans or photocopies) will there be a charge, and if so, how much? If network participants provide loans or copies will they be governed by broader institutional policies (i.e., University library policies) governing that service? Will the Solicitor General financially assist or underwrite costs for network participants providing document delivery? If so, will

such assistance be based on an outright grant or on charge-backs by network members for actual services rendered? Should materials be loaned or only copies made? Loans invite risk of loss of originals and should be avoided. What obligation does an inquirer have to try to locate materials himself in his own community, as opposed to expecting 100% backup for every request from the CRIMDOC network, which, although it would have the backup documents, shouldn't necessarily have to provide things available elsewhere?

There are basically two options for document delivery: make it a shared network responsibility or make it a centralized responsibility of the Solicitor General's library. The latter is the more efficient and consistent method, but inquirers of the data base should be encouraged to try to locate documents locally first before requesting them from Ottawa. Centralized document delivery is a labour intensive service, and should be adequately staffed. If document delivery is centralized then copies of all documents indexed and input in network locations must be forwarded for storage. This requires full cooperation.

A centralized document delivery service would have two basic options for storage of those documents: hard copy (i.e., print) or microform. The Price Waterhouse Associates study on Library automation (1981) recommended that all data base documents be microformed. This is a costly method which could involve purchase or lease of micro-reduction equipment suitable for the task and the training of staff to handle it, or the documents could be sub-contracted out to a company for that purpose. Time-lag should be compared for both methods. The advantage of microfiche copies of documents is their ease of storage, and "hard copy" photocopies could

be made on a reader-printer machine. Microfiche copies of all documents could increase the reliability of document delivery considerably. Print copies, as opposed to microfiche, could vary in size, shape, binding, etc., making storage and photocopying more difficult, although storage could likely work satisfactorily with accession-numbered vertical files.

A document delivery system must, above all, be efficient. At the time of data entry documents must be assigned a location with specific storage information. The accession number could be automatically generated upon data entry, and could be applied to either print copy or microfiche files. Location codes (i.e., network participants) could be changed but accession numbers maintained upon receipt of the documents at the Solicitor General's Library from the network location.

Staffing must be sufficient to allow for minimum turn-around time in filling requests. Staffing will be particularly important for the administrative records required to handle invoicing and any other paperwork resulting from the document delivery service.

Of concern in the area of document delivery is the problem of restrictions on the distribution of certain source documents. This could be handled by security control functions of the software which, upon accepting the access number of the person searching the file, would automatically open or restrict certain levels of the data base to that user. Another option would be to have a note in the availability field which advises the user to address inquiries concerning that document to a specific person or office. This could be a more sensitive security measure, as Freedom of Information Legislation might permit access to such material once identified.

To operate efficiently a document delivery system must have policies, must be adequately staffed and funded, and documents must be housed in such a way that they are easily stored and retrieved to reflect data base holdings.

Summary and Recommendations

This third report serves to reinforce the recommendations made in the first two reports on the matters of indexing and abstracting standards and vocabulary control. This report goes further, however, in highlighting the administrative considerations involved in establishing the data base as a whole. The Price Waterhouse study (1981) ascertained a clear need for a CRIMDOC data base, both by researchers within the Ministry (particularly the Secretariat) and librarians and researchers across the country. Now that further background work has been done, it is suggested that the following recommendations be given due consideration.

Recommendation #1: It is recommended that a committee of the Secretariat, including representatives of the Library, re-assess and confirm in an internal report, their objectives for the CRIMDOC data base, based on background information available to them from either this report (and its first two parts) or the Price Waterhouse Automation Feasibility Study.

This internal report can be quite succinct, and need only make reference to certain information if the committee is endorsing recommendations made in earlier reports. This report should record the official objectives of the data base.

Recommendation #2: Based on the above report and the official objectives of the CRIMDOC data base, the Solicitor General should begin specific planning and budgeting for the project.

As various components of creating the data base will be particularly time consuming (eg. creating and translating the indexing vocabulary), a schedule for the progress of the data base could be as follows: Fiscal Year 1983-84 - work on vocabulary (see CRIMDOC report #2); FY 1984-85 - create pilot data base and evaluate; FY 1985-86 - open up data base to full network participation; FY 1986-87 - evaluate whether or not CRIMDOC should be made commercially available. Budget planning should also begin immediately.

Recommendation #3: It is recommended that CRIMDOC be established as a private file, independent (from a systems perspective) from other library functions.

A private bibliographic file could be up and available to the Canadian criminal justice research community long before an integrated system could ever be available, and at less cost to the Solicitor General. The R.C.M.P. Library has chosen to establish an integrated system using STAIRS for cataloguing, reference inquiry and circulation records. At this point, with the Solicitor General's Library using UTLAS for cataloguing, and UTLAS not having the capacity to handle non-bibliographic items such as projects, programs and research in progress, it is more appropriate for the CRIMDOC data base to be built as an independent file, to be used as any other independent data base presently used in the Library.

Recommendation #4: It is recommended that work begin immediately on matters relating to acquisition of the NCJ Thesaurus tapes and the creation of the CRIMDOC vocabulary.

The creation of the vocabulary is going to be a very labour-intensive and costly process, but it is essential that this stage be begun immediately particularly since negotiations for the purchase of the NCJ Thesaurus and the decision regarding thesaurus software are interrelated.

Recommendation #5: It is recommended that a systems analyst write the technical specifications for the CRIMDOC data base and make recommendations on hardware and software best suited to the objectives of the data base.

It is not recommended that permanent data processing staff be appointed but rather that an analyst on a service contract assess technical systems specifications. Ongoing programming needs can be met by the service bureau on a fee-for-service bases. It is suggested that this person evaluate thesaurus software recommended in the 2nd CRIMDOC report on indexing vocabulary, namely LEX-INC. and LEXICO, as well as contacting Systemshouse, Inc., Ottawa, (re: ISIS/MINISIS), Infomart (re: BASIS), QL Systems (re: QL Search) and IBM (re: STAIRS) as possible sources of private file software and distributed system service bureaus. Some of the above have a thesaurus capability, such as Infomart's BASIS, which is now being used by the Ontario Ministry of Labour Library in Toronto for an article retrieval file and the Ontario Labour Relations Board Library for a case reports file.

Recommendation #6: It is recommended that, in the interests of inviting network participation, funds be budgeted for possible grants to those members, and that based on the official objectives for the data base, that various governmental and academic bodies be approached as possible participants.

A network for CRIMDOC holds the greatest promise of making a broader range of materials available in the data base, and appears to be the preference of the Solicitor General. There must, however, be a commitment on the part of the Ministry to continued and adequate financial support for the CRIMDOC data base before network participation is invited. While some potential network participants would be very interested in creating and maintaining such a network, university funding in most provinces is such that additional financial resources are necessary to become a participant. It would be beneficial to involve at least one network participant in any pilot project data base in order to adequately evaluate the effect of decentralized document selection, data entry and document delivery to the Ministry Library.

Recommendation #7: In order to provide for the creation and maintenance of the CRIMDOC data base, which will place new administrative, professional and clerical demands on the Library, it is recommended that additional staff be hired on both a permanent and temporary basis to execute those new duties.

While the Price Waterhouse Associates Automation Feasibility Study (1981: Appendix B-4) provides some estimates of staff and equipment requirements, which were based on very hypothetical expectations of the data base, there is some misrepresentation of time factors. It is estimated for example, that abstracting in a manual system could be done by professional staff at a rate of one item per hour and in an automated system at a rate of 100 per

hour. This is not the case. Indexing and abstracting are intellectual activities which do not change according to the system used. The storing of abstracting or indexing would be increased in an automated environment, but the rate of actual abstracting would remain at one per hour.

Staff positions likely to be affected and/or required could be as follows:

Criminologist/Reference Specialist: It is likely that this existing position should assume administrative responsibility for the data base; the time spent on the CRIMDOC data base in the first year or two would be significant, as it would involve work with computer hardware and software vendors, network participants, and the thesaurus construction committee.

Lexicographer: This should be a one-year contract position for the verifying of the base vocabulary and participation in the translation committee.

Linguist/Translator: It might be possible to second one person from the Secretary of State for a one year period, or else it will be necessary to hire a person for a one year contract.

Librarian/Indexer: One person will be required to do indexing and abstracting of documents on an on-going basis, and this should be a permanent position. This person could likely create the pilot project data base in approximately four months. Duties of this position, once established could also include selection and acquisition of documents, as well as the provision of searches and reference service.

Support Staff: A staff member with clerical and/or technical (i.e., library technician) skills will be required on an ongoing basis to enter indexing and abstracting at the terminal or word processor and to provide services such as retrieval, photocopying and mailing involved in document delivery.

Systems Analyst: See Recommendation #5 above.

During the course of preparing this background material for the Solicitor General many persons have expressed a keen interest in seeing the CRIMDOC data base come to fruition. It also appears that there would be a receptive commercial market for the data base several years from now if

the current trend in library use of on-line reference services continues, and if a good file is created. While the CRIMDOC data base does not aspire to be as large as the NCJRS data base, it does aim to serve the needs of the research and library community by keeping track of research and documents and avoiding unnecessary duplication of projects.

Appendix A:
Persons Contacted

Persons Contacted

The following persons were contacted in-person or on the telephone for information or opinions concerning hypothetical circumstances regarding a proposed criminal justice data base and/or for information on their own experience with information storage and retrieval systems. For their cooperation and assistance I am grateful. Any reference to specific individuals or systems are the interpretation of this author, and responsibility for those interpretations is that of this author alone.

Dr. Susan Painter, Head, National Clearinghouse on Family Violence,
Health and Welfare Canada

Ms. Paula Mitchell, The Register of Research and Researchers in the Social
Sciences, Social Sciences Computing Laboratory, University of
Western Ontario

Mr. Brian Morrell, Manager, Special Libraries, UTLAS (University of Toronto
Library Automation Systems)

Ms. Jean Benson, Account Supervisor, Infomart

Ms. Barbara Hopkins, Head Librarian, Ontario Labour Relations Board

Appendix B:

Footnotes

Footnotes

1. "Indexing and Abstracting Standards for a CRIM/DOC Reference Information Bank: Phase I Report: Bibliographic and Abstracting Standards." Prepared by Catherine J. Matthews, Centre of Criminology, University of Toronto, February 1982; and "Indexing Vocabulary for a Criminology Documentation Data Bank (CRIMDOC): Phase II Report." Prepared by Catherine J. Matthews, Centre of Criminology, University of Toronto, May 1982. (Solicitor General Contract No. 7270-29).
2. Price Waterhouse Associates, Management Consultants. "Ministry of the Solicitor General Criminology Documentation Centre Automation Feasibility Study." March 1981.
3. Marcus, Richard S. and J. Francis Reintjes (1981). "A translating computer interface for end-user operation of heterogeneous retrieval systems.1. Design", Journal of the American Society for Information Science 32 (July 1981):287-303.

Appendix C:
Bibliography

Bibliography

More comprehensive bibliographies of general background readings relating to the creation of a data base, and to matters of indexing and vocabulary control appear in the bibliographies accompanying the first two reports on the CRIMDOC data base.

Bourne, Charles P.

- 1980 "On-line systems: history, technology and economics." Journal of the American Society for Information Sciences 31(3, May 1980): 155-160.

Dolan, Donna R. and Heron, Carol E.

- 1979 "Criminal justice coverage in on-line databases." Database 2(March 1979):10-32.

Dowlin, Kenneth (Issue editor)

- 1981 "The electronic library." Drexel Library Quarterly 17 (4, Fall 1981):1-120

Grosch, Audrey N.

- 1979 Minicomputers in libraries 1979-80. White Plains, N.Y.: Knowledge Industry Publications Inc.

Harjee, Nimira

- 1982 "Interlibrary loan in an integrated environment." Special Libraries 73(1, January 1982):1-5.

Klempner, Irving M (Issue editor)

- 1981 "Information technology and special libraries." Special Libraries 72(2, April 1981):95-174.

Kruger, Donald R.

- 1981 "Issues and applications of microcomputers for libraries." Canadian Library Journal 38(October, 1981):281-285.

Kunkel, Douglas F.

- 1979 "So you want to build a network." In Clinic on Library Applications of Data Processing, University of Illinois at Urbana-Champaign, 1978. Problems and failures in library automation. Edited by F. Wilfrid Lancaster. Urbana-Champaign, Ill.: University of Illinois, Graduate School of Library Science. pp. 50-59.

Lancaster, F. Wilfrid, Drasgow, Laura S. and Marks, Ellen B.

- 1980 "The role of the library in an electronic society." In Clinic on Library Applications of Data Processing, University of Illinois at Urbana-Champaign, 1979. The Role of the Library

LIBRARY
MINISTRY OF THE SOLICITOR
GENERAL

MAR 22 1983

BIBLIOTHÈQUE
MINISTÈRE DU SOLICITEUR GÉNÉRAL

in an electronic society. Edited by F. Wilfrid Lancaster. Urbana-Champaign, Ill.: University of Illinois, Graduate School of Library Science. p. 162-191.

Levine, Gwen Reveley

- 1981 "Developing databases for on-line information retrieval." Online Review 5(2, 1981):109-120.

Marcus, Richard S. and Reintjes, J. Francis

- 1981 "A translating computer interface for end-user operation of heterogeneous retrieval systems. 1. Design." Journal of the American Society for Information Science 32(July 1981): 287-303.

Moureau, M.L.

- 1978 "Problems and pitfalls in setting up and operating an online information service." Online Review 2(3, September 1978): 237-244.

Sessions, Vivian S.

- 1971 URBANDOC/A bibliographic information system. Technical supplement 1. General manual. New York: The Graduate Division, City University of New York.

Swanson, Rowena Weiss

- 1975 "Design and evaluation of information systems." Annual Review of Information Science and Technology 10(1975):43-101.

Tsichritzis, D.C. and Lochovsky, F.H.

- 1978 "Designing the data base." Datamation (August 1978):147-151.

Uluakar, Tamer; Pierce, Anton R.; and Chachra, Vinod

- 1981 "Design principles for a comprehensive library system." Journal of Library Automation 14(June 1981):78-89.

Wall, Eugene

- 1978 Vocabulary management: a technical report. Rockville, MD: Aspen Systems Corporation.

Williams, Martha E.

- 1978 "Online retrieval-today and tomorrow." Online Review 2(4, December 1978):353-366.

- 1980 "Future directions for machine-readable data bases and their use." In Clinic on Library Applications of Data Processing, University of Illinois at Urbana-Champaign, 1979. The Role of the Library in an electronic society. Edited by F. Wilfrid Lancaster. Urbana-Champaign, Ill.: University of Illinois, Graduate School of Library Science. pp. 82-93.

SOL.GEN CANADA LIB/BIBLIO



000014454

LIBRARY
MINISTRY OF THE
ATTORNEY GENERAL
MAR 22 1983
BIBLIOTHÈQUE
MINISTÈRE DU SOLICITEUR GÉNÉRAL

Date Due

APR 14 1983

Z Matthews, Catherine J.
699.5
.C76 Indexing and abstract-
M3 ing standards for a
1982 CRIM/DOC reference
pt.3 information bank.

