

This item was submitted to [Loughborough's Research Repository](#) by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

A computerized finding aid for the Arnold Papers held in the archives of Rugby School

PLEASE CITE THE PUBLISHED VERSION

PUBLISHER

© Martyn F. Brawn

LICENCE

CC BY-NC-ND 4.0

REPOSITORY RECORD

Brawn, Martyn F.. 2019. "A Computerized Finding Aid for the Arnold Papers Held in the Archives of Rugby School". figshare. <https://hdl.handle.net/2134/10397>.

This item was submitted to Loughborough University as a Masters thesis by the author and is made available in the Institutional Repository (<https://dspace.lboro.ac.uk/>) under the following Creative Commons Licence conditions.



For the full text of this licence, please go to:
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

LOUGHBOROUGH
UNIVERSITY OF TECHNOLOGY
LIBRARY

AUTHOR/FILING TITLE

BRAWN, M.F.

ACCESSION/COPY NO.

040075 P76

VOL. NO.

CLASS MARK

T

0400758768



**A computerized finding aid for the Arnold Papers
held in the archives of Rugby School.**

by

Martyn F. Brawn B.A.

A Master's Dissertation, submitted in partial fulfilment
of the requirements for the award of the
Master of Arts Degree of the
Loughborough University of Technology.

September, 1993.

Supervisor: Professor M.A. Reed, M.A., L.L.B., Ph.D., F.R.Hists., F.S.A.
Department of Information and Library Studies

V 99 23135

Loughborough University	
of Technology	
Date	December 1993
Class	T
Acc.	
No.	040075 F76

Contents

List of figures	iv
Acknowledgements	v
Abstract	vi
Chapter One - Introduction	
1.1 Background	1
1.2 The Arnold Papers	4
1.3 Objectives	7
1.4 Outline of chapters	8
References	10
Chapter Two - The nature and theory of archival description	
2.2.1 Introduction	11
2.2 Moral defence of archives	12
2.3 Definition of terms	13
2.4 Arrangement	15
2.5 Description	18
2.5.1 Finding aids	22
2.6 Indexing	23
References	28
Chapter Three - The practice and application of archival description	
3.1 Existing arrangement of the Arnold Papers	30
3.2 Methodology	32
3.2.1 Data gathering	36
3.2.2 Data analysis	38
3.2.3 Description	38
3.2.4 Indexing	41
References	44
Chapter Four - Computerization of archival description	
4.1 Introduction	45
4.2 Evaluation of the software	46
References	49
Chapter Five - Development of the finding aid	
5.1 Introduction	50
5.2.1 Functional specification	51

5.2.2	Technical specification	52
5.3	Database design	55
5.4	Using the finding aid	58
5.5	Conclusion	71
	References	72

Chapter Six - Conclusions

6.1	Problems encountered	73
6.2	Successes and innovations	75
6.3	Recommendations	75
6.4	Conclusion	76

Appendix - The Arnold family tree	77
--	-----------

Bibliography	78
---------------------	-----------

List of figures

Figure 1	database design	56
Figure 2	query example 1	58
Figure 3	result of query example 1	59
Figure 4	query example 2	60
Figure 5	result of query example 2	61
Figure 6	query example 3	63
Figure 7	result of query example 3	64
Figure 8	query example 4	66
Figure 9	result of query example 4	67
Figure 10	query example 5	69
Figure 11	result of query example 5	69
Figure 12	ITEMS data entry screen	70

Acknowledgements

I would like to express my gratitude to Professor Michael Reed, my supervisor, for his help in arranging this project and for his helpful comments on the manuscript prior to its completion.

I am also most grateful to the librarian and archivist at Rugby School, Mrs Jenny Macrory, for her help in making available to me the Arnold Papers and for organizing accommodation for me on school premises. I am also grateful to the staff of the works department of Rugby School for all their considerable help in finding furniture for the flat provided by the school, and for their Herculean and persistent efforts which were necessary to get the bed into the flat.

I would like to thank my family for their patience, support and encouragement during the long weeks of my tapping away at my word processor, and Cliff Steele for all the help he provided me with by arranging for the printing of the finished manuscript.

Abstract

This dissertation is concerned with the provision of a computerized finding aid for the Arnold Papers held within the archives of Rugby School. The first chapter provides a brief overview of the school and the life and times of Dr Arnold. The objectives outlined in this chapter are the production of a useful and usable computerized finding aid which will provide an infrastructure to enable the description of the Rugby School archives to be expanded to other collections.

Chapter two is a review of the nature and theory of archival description as described through the writings of respected authorities in the field. Such concepts as moral defence of archives, arrangement, description and indexing are discussed.

The third chapter builds on the discussion in the previous chapter in terms of the actual circumstances prevailing in this project. The chapter deals with the practical application of the general principles of archival description to a specific collection. The inherited arrangement of the Arnold Papers is discussed along with the methodology employed for the data gathering, data analysis, description and indexing of the holdings.

A fourth chapter deals with the peculiarities of computerizing the finding aid, explaining the points of departure between traditional, paper-based, description techniques and those of electronic representation of the collection in a computer database. The software used for the project is evaluated in terms of the concepts discussed in chapters two and three.

The development of the finding aid follows, in chapter five, with the functional and technical specifications for the functionality of the finding aid and an explanation of the database design process. A section on using the finding aid for retrieving information from the collection is included in this chapter, supported by examples illustrated by screen-prints from the database itself.

The dissertation is concluded by a sixth chapter which pulls together the problems encountered during the project, its successes and innovations, and recommendations for future action in terms of expanding the Rugby School archival description project by means of the framework afforded by the Arnold Papers database.

Chapter One - Introduction

1.1 Background

In 1567 died the London grocer Lawrence Sheriff. Born the son of a Rugby yeoman, Sheriff rose to become purveyor of spices to Queen Elizabeth I. In his will Sheriff left £50 for building and a further £100 for maintaining a schoolhouse at Rugby. A subsequent codicil to the will changed the £100 to a bequest of one third of his Middlesex estates. This change of circumstances proved crucial for the long-term development of the school into one of the foremost public schools in the country, for the land in Middlesex was at what is now Holborn, in London. Worth some £8 per annum in the sixteenth century it was, by 1891, worth c. £5000 (1).

According to Sheriff's will the foundation was to be a free grammar school to provide education for the children of Rugby and nearby Brownsover. By the late eighteenth century the school had become known simply as Rugby School, by which name it is known today.

Dr Thomas Arnold, who in December 1827 was to begin his tenure as the most influential and, to his contemporaries, controversial headmaster of Rugby School, was born in the Isle of Wight in 1795. The son of William Arnold the Collector of Customs and Postmaster for the Isle of Wight and Collector of Dues for Trinity House, Thomas was brought up at the family home of Slatwoods, West Cowes. Thomas's education began at the small grammar school at Warminster where he followed in the footsteps of his elder brother Matthew who was to go on from Warminster to Oxford. The young Thomas Arnold displayed a precocious intelligence whilst at Warminster, so much so that by the time he turned thirteen the masters of Warminster were confessing the inadequacy of their establishment for the continuation of Thomas's education. In 1807 he was moved to Winchester College, from where in 1811 he went up to Corpus Christi College, Oxford, at the tender age of 16. Three years later Arnold graduated in the first class and against stiff opposition was elected to a fellowship of Oriel College. In 1818 Arnold was

ordained Deacon by the Bishop of Oxford and in 1820 he left Oriel to marry Mary Penrose of Fledborough, Nottinghamshire, the sister of his great friend Trevenen Penrose.

In order to support his new bride Arnold took to the occupation of schoolmastering. Following the example of his friend from Corpus days, John Buckland (who in 1816 had married Thomas's elder sister Frances) Arnold became a partner in Buckland's school at Laleham in Middlesex. Between 1821 and 1835 Mary Arnold gave birth to nine surviving offspring, four daughters and five sons. Amongst the latter was Matthew Arnold the poet, named for his uncle, Thomas's elder brother Matthew.

In the closing days of December 1827 Thomas Arnold was elected headmaster of Rugby School. This appointment surprised many, Arnold himself included. The new headmaster had youth, his reputation made from the conduct of Laleham old boys at Oxford and, most importantly, the patronage of his friend Edward Hawkins of Oriel to thank for swaying the Trustees' decision in his favour (2). From the time of his appointment to that of his death in June of 1842 Thomas Arnold was no stranger to controversy and unpopularity. He gained a widespread reputation as a political radical. His outspoken support of the Great Reform Bill of 1832 and apparent intellectual support of the French Revolution of 1830 combined with his position as moulder of young personalities made him a prime target for the myriad of conservative papers, magazines and reviews. As is still the case today the schoolmaster who speaks out on issues of politics invites the accusation of indoctrinating his pupils with his own radicalism. Arnold fuelled the sentiments against him with his writings on matters religious.

The 1830's, Arnold's first decade as headmaster of Rugby, coincided with the beginning of the Oxford Movement led by Henry (later Cardinal) Newman. The Newmanites, or 'Tractarians' after Newman's *Tracts for the times*, saw the supremacy of State over Church, or even an alliance of Church and State, as a threat to the theological autonomy of the Church. These 'New Catholics' or Apostolicals emphasized the role of the clergy, as the clergy were the living embodiment of the Church and the Church provided the only unbroken chain of authority from Christ down to the present

day, and, as such, was the only guarantor of the common and fundamental Christian beliefs.

The Newmanite view of Christianity was a political one. That is, when the intellectual and theological basis of belief and Church doctrine and ritual has been understood the human structures and hierarchies which perpetuate this intrinsic truth become paramount. Newman's division of Christianity was threefold: it had prophetic elements derived from Christ as prophet, it had sacerdotal elements derived from Christ as priest and it had ruling elements derived from Christ as king. These three elements led necessarily on to theology as the prophetic element, mystery and ritual as the sacerdotal element and the rigid and worldly hierarchical order of the priesthood as the ruling element.

Arnold's view of Christianity was a philosophical one. The great benefit of Christianity was its practical function of providing a moral and ethical code by which to lead one's life. Arnold distinguished between Christian philosophy - religion - and Christian polity - the Church. The former was successful and the latter corrupt and in urgent need of reformation.

Arnold despised what he termed 'priestcraft', the division of society into two camps: the clergy and the laity, with the clergy appropriating the power. His particular vehemence against these 'Frenchmen in redcoats', these Catholics within the Church of England, which found its outlet via the nib of his pen in several important pamphlets and articles during the 1830's, stemmed from the way the Newmanites seemed to be trying to narrow the participation of the laity just at the time when, it seemed to him, the Church should be reformed to widen lay participation.

Battle was joined between Arnold and the Newmanites in 1833. Newman himself dates the start of the Oxford Movement to John Keble's sermon in July of that year on 'National Apostasy'. Also in 1833 Arnold published his pamphlet on *The principles of Church reform* (3).

The intensity of the animosity between Arnold and the Newmanites and their supporters in the country at large is a very important aspect of his correspondence of

these years as well as of his reputation and progress as schoolmaster and cleric. This issue of Christianity; what it should mean and how it should manifest itself in the national Church, were the great questions of the day and it is against this background that the first five volumes of the Arnold Papers dealt with in this project must be approached.

1.2 The Arnold Papers

The types of records held in the Arnold Papers are quite narrow in their diversity. The overwhelming majority are letters (outgoing) from Arnold to his numerous correspondents. Many of these letters are photocopies of originals held in other repositories, such as Leeds University's Brotherton Library, the National Library of Scotland, Winchester College Library and the library of Oriel College, Oxford. The letters are chiefly from Thomas Arnold to family, friends and pupils' parents. Items discussed in the letters range from family trivia, through current affairs to great questions of politics, religion and the theory and practice of education which exercised the intellectual elite of the day.

Also included within the papers are books of sermons, diaries, printed circular letters, the manuscript of the tragedy Arnold wrote when he was just seven years old, entitled 'Piercy', and the newspaper in which the candidates for the headmastership, including Arnold, were announced in 1827.

Due to pressing demands of time and the decision to index comprehensively at individual item level only the first five of the nine volumes of miscellaneous papers have been covered by this current project. It was thought preferable to describe five volumes at a uniform level of detail than to go on to the remaining volumes at an ever decreasing level of detail. This latter course of action would have led to a misleading and inconsistent description and would have compromised the finding aid's efficiency. Further, as the project was undertaken partly to provide the Rugby School archives with the computerized infrastructure for archival description which could be expanded to include other collections within the repository, the paramount concern was to complete

the computerized framework upon which this subsequent expansion could take place and when the school was able.

The collection of letters, forming the bulk of the papers, contained no incoming correspondence with Arnold as the recipient. It is clear, therefore that the surviving letters were initially kept for sentimental reasons by the recipients. It is perhaps indicative of this process that we have such a long series of surviving letters written by Arnold to his Aunt Susan Delafield, beginning with Arnold at Warminster Grammar School and continuing through his time at Winchester and Oxford. The chance of which recipient was a hoarder of letters dictated which ones we have available to us today. As Arnold became older and more well-known in national circles letters may have been kept with a consciousness of their intrinsic interest as having been written by a historic personage. Many of the original manuscript letters in the collection have been mutilated to the extent that the signature of Arnold has been cut away. Clearly the autograph of this famous educationalist had, at some point in the letters' history, been of more interest to the holders than had the letters and their contents themselves.

The coming together of the Arnold Papers at Rugby is an instance of an artificial collection. Not an artificial collection in the sense of a collector or antiquarian accumulating old or thematically related material for its own sake, but of an institution collecting items related to a man intimately connected to it both in the public and institutional consciousness. There is, therefore, not the 'organic' element to this collection which writers on the theory of archive administration emphasize when promoting the understanding of the creating administration. In this case the creating administration is the pen of Thomas Arnold. There are no committees or sub-groups to be borne in mind.

Successive librarians at Rugby School have contributed to the accretion of material to this collection for three principal reasons. Firstly, the librarian receives, daily, several written and telephoned enquiries from all around the world, many relating to Arnold and/or Rugby at the time of Arnold. The School is, for the majority of enquirers, the natural focus for their researches. The accumulation of papers, including photocopied

surrogates from other repositories, allows for the centralization of primary source information on Arnold at the place where it is most frequently needed.

Secondly, for purposes of completeness and institutional pride in a renowned former headmaster a collection of papers relating to Arnold has, where possible, been enhanced with the imported surrogates. And, finally, donations from individuals to the school have been accepted in the spirit in which they have been made. Over the years many items of limited interest or value to the school have been gracefully accepted by the librarian and, as often as not the grace of their acceptance has been rewarded by further donations of important, valuable and unique documents.

The archives of Rugby School as they currently stand are exploited merely by the memory, experience and hard work of the current librarian. Room is short, many collections lie untouched in ribboned bundles inside cardboard boxes. Knowledge of their contents is, necessarily, haphazard and chancy. Answers to written or verbal enquiries are reliant upon the librarian's memory and/or educated guesswork as to where to find the relevant information. We cannot, therefore, say with any honesty that the archives of Rugby School are systematically exploited or even exploitable. Without logical arrangement, description and indexing the collections in the school archives will continue to hide their secret knowledge as they have done since their creation. This situation, then, is the one which demands that some form of framework or infrastructure be put into place so that the logical arrangement, description and subsequent exploitation of the school's archival resources can be undertaken.

The Arnold Papers were chosen to blaze the trail of this effort because they are the most frequently consulted documents in the school archives. A finding aid for this collection would be most immediately useful in terms of answering enquiries as well as providing the material for an archival description infrastructure as adequately as any other collection in the school.

1.3 Objectives

The objectives of this archival description project are fivefold: firstly, the primary aim is to produce a useful and usable finding aid for the Rugby School librarian to use in her day to day activities of controlling the archives and answering researchers' enquiries.

The second aim is to provide the school with a foundation of a computerized system for the management, control and exploitation of the archives in its care. A carefully designed and implemented database should enable the expansion of the descriptive process to the rest of the Arnold Papers and to other collections to go ahead more easily and more quickly than would be the case if the task had to be tackled from scratch.

The third objective of this project is the realization of the advantages for data manipulation, within the finding aid at the time of searching, of a computerized solution. Manual finding aids are, by their very nature, inflexible things. When designing a manual finding aid it is necessary to make judgements as to the likely needs of the user of the system. The arrangement of the data in a single way is also unavoidable. With a computerized finding aid the data within it can be sorted and arranged according to the criteria of the researcher. For example, sorting by date and/or person allows for more precisely directed search strategies. So, regardless of the type of enquiry being made the researcher is not restricted by the preconceptions of the archivist and has a much increased choice of access point to the information the records contain.

The fourth purpose of this project is to index the five volumes of the Arnold Papers being dealt with comprehensively on persons, places and subjects/concepts down to an individual item level. In this way the maximum flexibility is maintained for the exploitation of the archives, in the sense of providing the most direct access point(s) for enquirers to the relevant document(s). Also, comprehensive indexing avoids the dangerous need to make judgements as to the persons, places or subjects which a researcher might find important. This is always a difficult path for archivists to tread for who can infallibly predict what areas may be of interest to future researchers? In an institutional archive with pressure of time and resource the answer is often that a

compromise has to be reached and some form of judgement exercised in order to provide the best level of access possible under the circumstances. In this case, however, where the time and resource is very limited a decision to deal comprehensively with a small subsection of the whole is permissible; especially so as a major objective of the project is to provide an infrastructure which may be followed in the future. In the case of letters, which make up the lion's share of the collection, a comprehensive index is justifiable even under more normal circumstances by the nature of the random and disconnected information they contain. A finding aid listing merely names of sender and recipient and date of writing goes little way towards providing the researcher with an idea of what information the items may contain.

The final objective of the project, and one which is implicit in the objectives which have gone before, is to provide the librarian, pupils and authorized researchers at Rugby School with a flexible system which will stand as finding aid, record control tool and model for expansion for some time to come.

1.4 Outline of chapters.

It only remains in this introductory chapter to give an outline of the chapters to follow so as to provide the reader with the context in which the thesis will be presented. The first chapter proper (i.e. chapter two) will discuss the nature and theory of archival description; the concept of the moral defence of archives, arrangement, description, finding aids and indexing. Chapter three will examine the existing arrangement of the Arnold Papers and discuss the methodology used in the creation of the finding aid. This chapter will be concluded by an explanation of the policy followed in reconciling the already explored theory of archival description and existing arrangement with the objectives of the Arnold Papers project, with reference to gathering and analysis of data, description and indexing.

The fourth chapter will deal with the theory of computerization of archival finding aids. The software to be used will be evaluated with regards to its suitability for archival description.

The general development of the database will be dealt with in chapter five. Sections on the functional and technical specifications will be included, as will an explanation of the design philosophy and of the use of the finding aid with example queries, forms, data input screens etc. Chapter six will conclude the dissertation with a review of the problems encountered, successes and innovations and recommendations for the expansion of the Rugby School archival description project.

References - Chapter One

1. Anderson, J. *Rugby School archives: The Arnold papers*. Unpublished M.A. dissertation, University of Loughborough, 1988.
2. Bamford, T.W. *Thomas Arnold*. The Cresset Press, London, 1960.
3. Trevor, Meriol. *The Arnolds*. The Bodley Head, London, 1973.

Chapter Two - The nature and theory of archival description

2.1 Introduction

The impact of high technology has increased the pressure, which has long existed, to structuralize the archival craft along the lines of 'library science' (1). Librarians, however, have care of books; volumes of systematized knowledge brought together consciously on a specific topic or set of topics. The pages which constitute a book have a carefully designed sequence and rigidly defined relationship to the pages around them. The binding of a book makes it a self-contained repository of knowledge which lends itself to generic schemes of classification, and, subsequently, the development of a 'library science'.

Archives are, by contrast, gatherings of materials at one remove from the unselfconscious and individual acts of creation of the various elements which go to make them up. Even archives which, at the time of their creation, are earmarked for permanent retention, such as the public records of the departments of central government, are primarily collections of functional documents. They have a working life for which they are designed. For the typical document the end of its working life coincides with the end of its actual life. Those documents conscripted into a second career as part of an archive group are largely in this happy situation through the good offices of random chance. From this premise it follows that the information contained within such agglomerations of documents does not lend itself to rigid schemes of classification in the same way that collections of books do.

Archivists must, therefore, beware the temptation of fitting the arrangement and description of an archive group into a pre-conceived system. Manuals containing the templates of such systems must be exploited selectively, as dictated by the circumstances of the repository and the requirements of the archive group in question.

2.2 Moral defence of archives

Having said that the archivist's art is just that; an art, not a science, it must be stressed that there are fundamental guiding principles which define the arena in which the vast majority of the archivist's work directed at making the information they contain accessible will take place.

Perhaps the most important of these fundamental principles is that known as the moral defence of the archives (2). The archivist has two main functions: to preserve physically the items in his care, and to make available the information they contain. Authorities differ as to the relative importance of these twin pillars of the archivist's role; Jenkinson, for example, writing in the 1930's (3) was firmly of the opinion that the primary task of the archivist was to provide a physical safe haven for the archives and only as a secondary duty, where such could be carried out without compromising the primary duty, should he concern himself with description and provision for researchers. In more recent years the exploitation of archival collections has become more and more important, especially to the fund providers. As a consequence the duty of moral defence of the archives has grown in relative importance in parallel with the drive to make archival collections yield up their information.

The first rule of moral defence is that collections of documents should not be split up. Enclosures should not be removed from letters, for example. For it is often the juxtaposition of related items which provides evidence for the provenance and authenticity of a particular document or set of documents within an archival series. Once a collection is split up the relevance, and knowledge of, their original relationship to one another is often lost with little or no hope of redemption. Information on context and background is often necessary to interpret the information contained within the archive(4).

The moral defence of an archive also implies the responsibility for the integrity of the collections concerned. That is, that no evidence be lost, and no evidence be

manufactured (5). The collections in an archive may be in one of four states of organization when an archivist comes to exercise his intellectual control over them:

- (a). The documents could be already arranged and a list provided.
- (b). The documents could be already arranged and no list provided.
- (c). The documents could have had no interference at all.
- (d). The documents might have been 'inherited' from an archivist who has already dealt with them.

In the case of (a) or (b) the arrangement should be checked independently. When no management information is provided, as in the case of (c) the archivist should investigate the structure of the creating administration and arrange the archives in such a way as to reflect this. In the case where a collection is 'inherited' from another archivist then two courses of action are available. Firstly the decision may be taken to re-arrange and, if thought necessary, to treat the collection as a new acquisition. Alternatively, the arrangement may be accepted. The purpose of archival arrangement, in Jenkinson's words, is to:

"marshal them in such a way that the archive significance of every document - its own nature and its relation to its neighbours - is brought out as clearly as possible (6)."

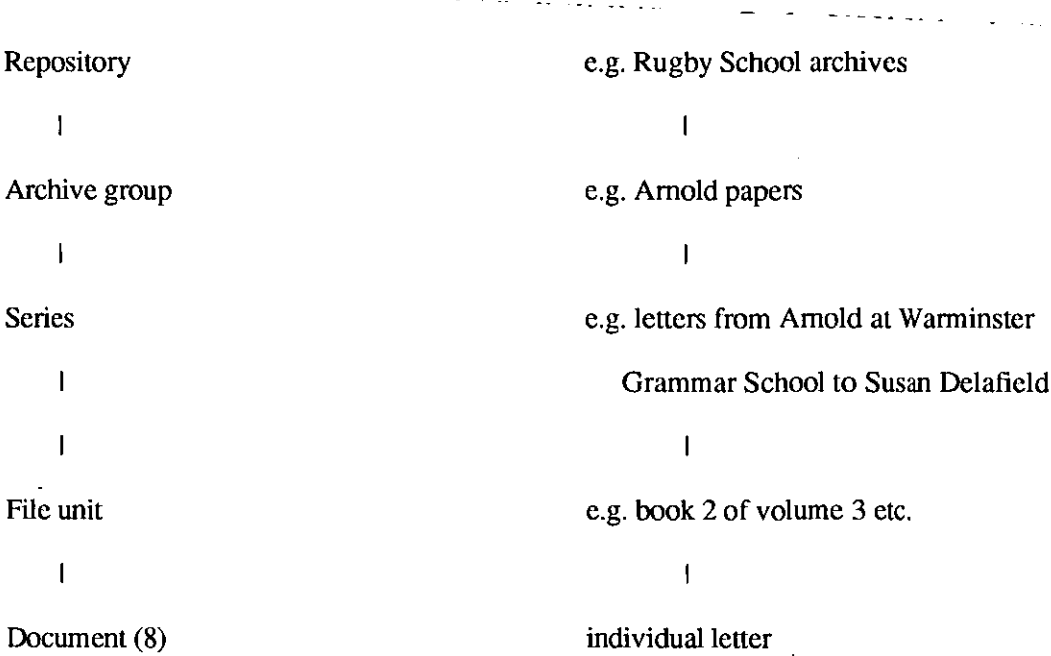
2.3 Definition of terms

The moral defence of an archive calls for the *arrangement* of the items within an *archive group*, the *description* of the group and the provision of a *list*, *inventory* or *finding aid* for the group. Before going further I shall provide some definitions of the words in italics.

The archive group is the largest body of related records organized within a particular repository, based on their relationship with one another. They may have been

created by the same person, by the same organization, or by a subset of an organization or a group of individuals (7). In terms of the Rugby School archives the collection of the papers of Thomas Arnold dealt with in this project constitute an archive group; as do similar collections of Vaughan papers, Bloxham papers etc., also at Rugby.

These archive groups together form the archive. The hierarchical relationship of repository down to individual item can be represented diagrammatically:



Arrangement, in archival terms, is the physical process which stems from the intellectual analysis and organization of the documents in the archive group. It is similar to the library process of cataloguing, but is done in relation to other archive groups, and by relations of documents with one another internally to the group, rather than by a system template such as the Dewey decimal system used in many libraries in this country(9).

Description of archival holdings is synonymous with any work carried out to prepare finding aids; that is, to facilitate extraction of information from the physical media of the archives (10). Whereas arrangement has an obvious physical manifestation, both arrangement and description are primarily intellectual processes. The physical manifestation of description is a finding aid of some sort. The most basic finding aid for

an archive or archive group is the inventory or list. As the term list implies, this is a listing of the contents of an archive or archive group.

2.4 Arrangement

Archival arrangement is a term used to cover the physical processes of archival organization. This is not, however, meant to suggest that arrangement is a mere mechanical task with no intellectual input. Far from it. Much energy is devoted by archivists to the arrangement of collections before they can be described. The subtle difference of emphasis between arrangement and description is this: the main product of arrangement is the layout and location of individual records and groups in relation to one another which is achieved through intellectual effort. The primary product of the description is the summary of the archive's holdings. This is an intellectual product, made physical only to enable its communication.

The requirement for archives to be arranged is summed up in the Society of Archivists' handbook: *Preparation of finding aids* :

"Most documents are part of a sequence, great or small, concerned with a particular event or series of events, a particular action or series of actions, and can only be understood and their information evaluated alongside their fellows (11)."

The evidential value of a collection is provided by its coherence. The name given to this concept is the sanctity of the archive group. To take as an example a series of letters comprising one side of an on-going correspondence, such a group of documents would almost certainly prove more valuable to a researcher as a coherent whole than would the sum of the constituent parts were they scattered, or even out of order. Allusions in a single letter might be made explicit in the next. Persons discussed by Christian names or occupational titles alone may be able to be fully discovered in a series of letters. The possibilities for interpretation of such a series arranged chronologically in groups by sender and recipient are very great when compared with the problems one

would encounter trying to evaluate the same letters scattered about the repository in the locations to which the vicissitudes of their active lives have condemned them.

As well as underwriting the evidential value of the information contained within the archives, arrangement is important for the understanding of the workings of the body which created them. If by His works shall ye know Him, then by its archives shall ye know an organization.

-An archive group arranged according to the model of the creating body allows the significance of the documents themselves to be more easily grasped. For instance, in an organization comprised of departments and committees and sub-committees etc. the resultant records should be grouped according to the department or committee which produced them. If the records of each such body are grouped in this way then by knowing the structure of the organization, which is captured for the student in the archival arrangement itself, then the machinations of the organization as a whole can be reconstructed and understood in a way which would not be possible were the proceedings of the individual bodies within the organization scattered at random around the archive. An arrangement arbitrarily by date, or by size or by material, or whatever, is just as obfuscating as no arrangement at all, because the logic of information flow which would otherwise provide the student with an Ariadne's thread would not be apparent. A large archive is a labyrinth; to find one's way one needs a map (a finding aid) and/or a logical path to follow. In labyrinths one always turns left; in archives one follows the structure of the creating administration.

When the creating administration is not an organized group of individuals but a single individual, such as Dr Thomas Arnold, the principle of arrangement still holds true. The individual may be thought of as a creating administration. His correspondence can only be understood in terms of when and to or from whom it was sent, and as previously stated can be more easily interpreted when arranged chronologically. An individual's papers, however, are, by their nature, more readily reconciled with the workings of the creating administration than are an organization's. This is because the documentary outpourings of an individual, even the most prolific individual, will not

reach the Byzantine standards of complexity of those of even the simplest organization with many members. Organization of individuals into structured groups necessitates communication. Communication requires creation of written records. Written records of today become the stuff of archives tomorrow.

So, the arrangement of the papers of an individual need not be as rigidly adherent to the principle of the sanctity of the original order as those of a large organization. As correspondence often constitutes a large proportion of the papers of an individual, as certainly is the case with the Arnold Papers, an arrangement which places letters in chronological order grouped by sender and recipient is a simple and effective form of arrangement. It also fulfils the principle of reflecting the structure of the creating agency because it preserves the correspondence as an organic whole. Before the age of telephone communication much correspondence was just that, correspondence. Two people exchanging letters over a period of time during which ideas and opinions are also exchanged, modified, built on and developed. A correspondence, then, can only be considered whole when comprising all the letters which went to make it up. The letters through which the development of ideas and thoughts can be followed. Individual letters, or letters to different recipients which happen to have been written contemporaneously, do not reflect this aspect of a series of correspondence.

In the case of the Arnold Papers we have only the side of the correspondence written by Dr Arnold. Unfortunately he does not seem to have placed as much importance in the retention of his own mail as some of his own family and friends obviously did. Nonetheless, one side of a correspondence still has many of the features regarding continuity and development with time as does a complete correspondence. So, arrangements of letters by date of writing alone is not enough. They must be grouped chronologically by sender and recipient in real or 'logical' bundles.

2.5 Description

Archives must be described, for only through description can control be exerted, and only controlled collections can be preserved and exploited (12). The term 'description' can be taken to encompass any work undertaken to prepare a finding aid (13). Finding aids are vital to the control and exploitation of archival holdings. Rarely are such holdings available for the public to browse as are mainstream library collections. A facility for browsing must be provided by use of a surrogate; a list or catalogue which gives basic information about the contents of the archive.

It is not only to provide users with the opportunity of 'viewing' the contents of the archives that a description of some type is essential but for the purposes of retrieval. As individual documents and files within an archive are not classified and catalogued, in the sense that library holdings are catalogued, they must be identified individually with unique identifiers, or reference numbers. Archival holdings are often shelved inside uniform boxes and, as such, are indistinguishable from one another whilst on the shelves. The system for providing each item with a unique identifier which ties it in with the box or bundle to which it belongs and, in turn, a reference system which locates that box or bundle unambiguously on the shelves is the bare minimum of description necessary if an archive is to be used. Without a rigorous system of locating items, then, they are as good as lost, even though they are on the repository's shelves somewhere.

The overriding consideration governing the choice of finding aid and level of description is the necessity of giving attention to the needs of the users and the circumstances of the individual repository. The 'rules' of archival description are not cast in tablets of stone, but are guidelines from which one may select or reject according to the situation. The important thing about archival description theory is that it gives the archivist an *aide memoire*, a checklist for the intellectual activity of analyzing the holdings and providing an aid to their exploitation.

An important aspect of archival description is the setting in context of the different groups within the repository. It is not possible, or even desirable, to describe the holdings

of anything but the tiniest and most homogenous archive as one whole. The holdings of the archives must be split down logically from the archive as a whole to individual documents or production boxes or bundles. As there is inevitably only one repository but many individual records the splitting of the archive for descriptive purposes must necessarily be a hierarchical split. A recent work by Michael Cook aimed at the standardization of archival description suggests the following template for describing holdings at different hierarchical levels:

<u>Levels</u>	<u>Name of descriptive field</u>
0	Repository
1	Management Group (Archive Group)
2	Group/Collection
2.5	Subgroup
3	Class
4	Item
5	Piece (14)

Within this framework the individual archivist must select the appropriate levels and detail to use bearing in mind the size, composition and projected use of the archives in his care. Also, any precedent set within the repository must be considered so that new work is compatible with and complementary to existing descriptions.

As well as deciding on which levels of description to use a decision has to be made regarding the detail at which each chosen level will be dealt with. This may range from a very high level description giving just an overview in text of the scope of the collection with inclusive dates, to a comprehensive index at individual item level, or any combination in between. The types of level used are dictated largely by the type of

collection being described. This also has an influence on the detail at which each level of description is carried out. The main factor, however, is the information one wishes to convey, which in turn is dictated by the use one envisages for the archive.

As an example of the way in which a traditional, paper-based, archival description might deal with a collection of documents, the Arnold Papers might be described at the following levels:

0	Repository	i.e. the Rugby School archives
2	Group/collection	i.e. the Arnold Papers
4	Item	i.e. individual books within volumes *
5	Piece	i.e. individual letters.

* The papers are mounted in acid-free booklets which in turn are held in volumes. Each volume contains, typically, some five booklets. For instance, volume 3, book 2 would be considered an item, in Cook's sense of the term, as it is an example of the smallest subsection of the collection which would be produced for a researcher.

The management group (level 1) concept is not relevant to a repository the size of Rugby School. Accruals are irregular and infrequent, so differentiations such as deposited archives, or department archives etc. are meaningless in this context. A number of collections of individuals' papers are held at Rugby, so description at the collection level makes sense, to differentiate between the Arnold papers and the papers of Vaughan and Bloxham, for example. The next relevant level is the booklet in which the individual papers are mounted and produced for enquirers. Finally the individual letters and documents might be included as the emphasis in a small collection should be on specific information held within individual items. In a large repository this level of description would not be possible due to constraints of time and the demands of other collections. In such a large repository the collections may well be described at a high level, by box or

bundle for instance, as a compromise between available resources and the need to provide for information retrieval.

The one to many relationship which characterizes the hierarchical levels of archival description are 'tied' together through the medium of the reference code. Using the Arnold Papers example already outlined, reference codes could be built up for each level in this way:

- | | | |
|---|-------|---|
| 0 | RS | denotes holdings belonging to Rugby School. |
| 2 | AP | would identify the documents as belonging to the Arnold Papers. |
| 4 | 1/1 | signifying volume 1, book 1. |
| 5 | 1...n | signifying item 1 in that volume. |

So, an individual item designated RS/AP/3/4/23 could be identified as the twenty third document in book four of volume three of the Arnold Papers held in the archives of Rugby School. This unique identifying code gives a 'name' to each document in the archive, allowing location control and citation. It also facilitates placement within the hierarchy of the archive. A reference code such as the one above clearly denotes an individual piece.

RS/AP/3/4/23 would be one of at least 23 pieces in the item RS/AP/3/4. Book four would be one of five books in volume three (RS/AP/3). This project has dealt with five of the volumes of the Arnold Papers, RS/AP. A traditional paper-based description would give entries of ever increasing detail at each of these levels, from collection provenance, custodial history, inclusive dates etc. down to individual indexing of subjects, people and places at piece level.

Whatever level of description is chosen the templates of respected authorities are not the final arbiters of action. All activities of archival description must be undertaken with the user's circumstances and needs foremost in one's considerations:

"In principle, the appropriate depth of description is determined primarily by its effectiveness in providing a representation of the original which will work correctly in the circumstances envisaged....(15)."

Again, whatever level of description chosen the product of any description is a finding aid. It is this finding aid which, ultimately, will determine the worth of the descriptive process.

2.5.1. Finding Aids

There are, basically, three types of finding aid:

1. those created for internal control of collections.
2. those produced for in-house reference.
3. those published for public use. (16).

A single finding aid may, of course, fulfil one, two or all three of these functions. The one prepared by this project for the Arnold Papers is designed to cover the first two of these uses. Gracy's description of a finding aid for use as an in-house reference tool sums up well the design philosophy of the Arnold Papers' computerized finding aid:

"this type of finding aid is analogous to a computer data bank, which a person can probe and search at length to extract the information (or in this case knowledge of the information) he desires (17)."

The finding aid is the physical manifestation of the analysis carried out during the description. The essence of a finding aid is provision for the uniting of enquirer with information held within the archive. This means not only identifying the document or documents appropriate to the search but locating them on the shelves and allowing for accurate replacement.

2.6 Indexing

An index is an access point to the information contained within an archive. It should provide a guide for consulting the catalogue, where there is one, or form a part of an integrated finding aid where concepts of separation into catalogue and index are not appropriate (18).

The most commonly used indexes are:

1. Place names
2. Personal names
3. Subjects (19).

The level of detail one provides in an index to a catalogue, or even if one is provided at all, is a matter for the judgement of the individual archivist:

"Some archivists have held that to approach any problem through a subject index is unsound, because the 'proper' route is by an analysis of the administrative machinery of the organisation which produced the document (20)."

This approach is, however, too dogmatic. It eschews the principle of flexibility and individuality of circumstances. Some archival situations will call for detailed subject indexes, some for only generalized subject indexes and some for none at all. Assuming that an index is deemed appropriate for the archive and collection in question a decision has to be taken before work commences at what level the index will be produced. On the assumption that the data gathering for the finding aid will be carried out once and once only then all requisite information must be gathered at that time. Generally speaking the smaller and/or more specialized the archive the greater the depth of indexing. As the size of the archive increases so does the necessity for making selections as to what to include

and what to exclude. It is imperative that, as far as is possible, the archivist keeps an open mind and does not 'negatively select' items to be left out by too rigid an adherence to a pre-conceived plan of what enquires will be made of the archive and what lines of study are legitimate and which are not.

As the Society of Archivists' handbook on the creation of finding aids warns:

"Anyone who has spent any time supervising a reading room will have encountered readers with well chosen and legitimate spheres of interest which could not have been foreseen by an archivist listing or indexing its documents (21)."

It is with this regard for an open-minded approach to the documents that the current project has been undertaken. By indexing the Arnold Papers comprehensively my intention was to avoid assuming areas of study to which particular documents may prove useful, and to work outwards from the documents themselves by indexing without making arbitrary judgements as to the worth of documents for some assumed area of study. For to do this one is, consciously or unconsciously, restricting the legitimate areas of research for which the papers might be used by one's own limited knowledge, experience and imagination. Better, then, to document what is there as clearly as possible and in such a way as to facilitate retrieval of the information relevant to any request if it exists, leaving judgements as to the worth of the information to the enquirer.

There are cases, though, where areas of enquiry which will prove popular can be identified and efforts be made to make sure that information relevant to those areas are easily accessible through different access points within the index. For instance, from past experience and from common sense, it is known that the Arnold Papers are frequently accessed to answer questions on the biography of Arnold, his friends and relations and on the history of the school and education in general. It is important to bear these things in mind when constructing the index. They may well be useful as a yardstick next to which

one can measure the retrieval performance of the index on less well-known or more obscure items.

Along with a warning against pre-conceived ideas on what enquirers will require of an archive one must be careful when professing comprehensiveness for one's index. For although my aim as regards the indexing of the Arnold Papers was to provide a comprehensive coverage of people, places and subjects it is inevitable that some items and some lines of enquiry will have been overlooked. An index must, therefore, tread the ground between being misleading through arbitrary selection made on the archivist's preconceptions and being misleading by pronouncing an index comprehensive and so consigning the inevitably overlooked items to the obscuring shadow of the illuminated bulk of the collection (22).

Most authorities agree that where indexing is considered appropriate in an archival finding aid three types of entry can be made to cover most situations, namely persons, places and subjects. The level of detail undertaken when indexing archives:

"Depends on the nature and scope of the archives being indexed, their rate of accumulation, and, of course, on your resources of time and staff, which means ultimately, money. Generally, one may say {1} the more detail the better; {2} an item not indexed is as good as lost; {3} once done however badly, indexing is unlikely to be repeated; {4} what seems unimportant to you *may* be just the clue the researcher is looking for - that is, we must, as far as possible, be customer- or user-orientated. (Also a large accumulation of individually minor references may build up to a meaningful picture.). All indexing involves subjectivity, and we must be aware of the dangers of this, but they can to a great extent be mitigated by thoroughness and an awareness of the uses, likely and unlikely, which will be made of our finished product (23)."

I have quoted this passage from Storey at some length because it not only underscores what I have written regarding the dangers of subjectivity but also because of the way in which it emphasizes the paramountcy of the user's need to retrieve

information. Whether the index is merely a secondary reference tool acting as a guide to the catalogue, which some authorities insist is the proper function of the index (24), or as part of an integrated finding aid and records management system, or provides the primary access point to the archives in terms of information and location, is, when compared with the obligation to fulfil the user's need, of secondary importance. Like all the principles of archive administration the indexing has a set of guiding wisdom. In the end, though, it is the archivist's responsibility to select and reject from this inherited wisdom according to the circumstances he finds himself confronted with. There is, of course, no substitute for experience in exercising this judgement. The guidelines handed down through the writings of experienced archivists should not be laid aside, therefore, without thinking through the implications.

It is on this note, then, that I quote again at length from Storey, with ten points to bear in mind when indexing archives; ten points to which I shall return in due course when discussing the design of the Arnold Papers database index in the next chapter:

- "1. Index in as much detail as resources (money, manpower, time) allow.
2. Establish criteria and institute classification schemes and guides at the beginning of the work.
3. Choose a system with some flexibility to allow for expansion or alteration.
4. Allow one slip or card for each entry.
5. Use multiple entry, except for 'built-in' cross-references such as ENGINEERING *see also* TRANSPORT.
6. Bring out as many proper names as possible.
7. Make the reference system quite unambiguous: distinguish between references to the list, for example to a page number, and references to document numbers.
8. Start an index in good time: your archive may as yet be small

and your memory excellent, but no one is infallible, and all archives tend to grow.

9. Remember that indexes can, and should, become works of reference in their own right.
10. Consider indexing not as doing the work of the researcher for him, but as merely making it possible for him to do his own work. (25)."

Overview inventories of correspondence will not reveal a great deal about the content of the letters, short of clues derived from the date of composition and name of originator and recipient which may be incomplete, misleading or even missing completely. For an archive collection consisting almost entirely of letters, as is the case with the Arnold Papers, there is ample need of a detailed index. As Storey says in his point number nine above, indexes can and should become works of reference in their own right. It is in keeping with the principles of flexibility and providing an integrated finding aid with access points at various levels of specificity that the Arnold Papers finding aid is designed around a detailed index. The indexes of persons, places and subjects form the core level of entry to the collection. The higher level of information such as custodial history, materials, size etc. very much go to enhance the utility of the indexes. It is the various people and things discussed within a letter which gives it its significance. That is why the detailed index is given the position of primacy in the Arnold Papers finding aid.

References - Chapter Two

1. Taylor, Hugh A. *The arrangement and description of archives*. International Council on Archives, ICA handbooks series, volume 2, London, 1980.
2. Jenkinson, Hilary. *A manual of archive administration*. Percy Lund, London, 1937.
3. *ibid.*
4. Cook, Michael. *Information management and archival data*. Library Association, London, 1993.
5. Jenkinson, ref. 2, p. 83.
6. *ibid.*, p. 97.
7. Cook, Michael and Margaret Procter. *Manual of archival description*. 2nd. ed., British Library, London, 1989, p. 15.
8. Gracy II, David B. *Archives and manuscripts: Arrangement and description*. Society of American Archivists, Chicago, 1977.
9. Cook, ref. 7, p. 257.
10. *ibid.*, p. 260.
11. Society of Archivists. *Preparation of finding aids*. Scheme of study, lessons 1-3, p.5.

12. Taylor, ref. 1, p. 15.

13. Cook, ref. 7, p. 260.

14. *ibid.*, p. 52.

15. Cook, Michael, quoted in: Miranda H. Scrase. *Creating an inventory: A finding aid to the records of the Derbyshire branch of the National Farmers' Union*. Unpublished M.A. dissertation, Loughborough University of Technology, 1991.

16. Gracy, ref. 8, p. 19.

17. *ibid.*, p. 30.

18. Society of Archivists, ref. 11, lesson 8, p. 7.

19. *ibid.*

20. *ibid.*, lesson 9, p. 11.

21. *ibid.*, lesson 8, p. 8.

22. Taylor, ref. 1, p. 16.

23. Storey, Richard A. 'Indexing archives'. *Indexer*. Vol. 5, no. 4, autumn 1967, p. 162.

24. Society of Archivists, ref. 11, lesson 9, p. 15.

25. Storey, ref. 23, p. 168.

Chapter Three - The practice and application of archival description

3.1 Existing arrangement of the Arnold Papers

The documents which make up the Arnold Papers are mounted in acid-free archival booklets, which in turn fit inside rigid volumes for storage on the shelves. The Arnold Papers are the only collection within the Rugby School archives that have been arranged or mounted, albeit rudimentarily, as they are the collection which is most frequently used and the collection which is most valued by the school.

There is no accompanying explanation of the methodology used for the arrangement, and it is not readily discernible from the arrangement itself. Book one of volume one contains just a single photocopied letter from Thomas Arnold's sister Susanna, dated sometime in 1799. Book two contains the original parchment certificates of ordination of Thomas Arnold as Deacon and to the priesthood respectively, and the licence to officiate as headmaster of Rugby School. Book three contains photocopies of letters, the originals of which are housed in the library of Winchester College. The first 15, written by Thomas Arnold to his aunt Susan Delafield from his school at Warminster, are arranged chronologically from 1803 to 1809. The next two letters; from Arnold to his mother, are also arranged chronologically, 18 November 1806 to 17 February 1807. This book is concluded by two single letters dated 1807 and 1809 respectively. Book four of volume one contains the original pencil manuscript of the tragedy 'Piercy' written by Arnold when he was seven years old. Book five contains two original poems, one by Arnold and one by his wife Mary.

I have listed at some length the contents of volume one so as to demonstrate the arrangement of the pieces that make it up. All the books, apart from book three, are very sparsely populated, keeping like documents with like documents. Why book one contains a single photocopied letter whilst book three contains 19 photocopied letters is unclear. The letters in book three are all copies of originals in Winchester College, so it is quite possible that they were all received at Rugby in one parcel and were arranged accordingly

by the principle of provenance. As letters they are rightly arranged chronologically within the series of sender and recipient. That is, the 15 letters Arnold sent to his aunt from Warminster Grammar School that are included in this book are arranged one after the other from the earliest to the latest. These are then followed by the two letters from Arnold to his mother, arranged chronologically in relation to one another but not in relation to the 15 previous letters. That is a sensible and consistent manner of arranging letters.

There are, however, many inconsistencies and/or mistakes. In volume two book three, for example, there are 14 photocopied letters from Arnold to his wife Mary. They are ostensibly arranged chronologically from the earliest dated 7 July 1823 to 15 September 1829, but AP/2/3/13 is dated 15 September 1829 whilst the next item, AP/2/3/14, is dated 17 April 1829. Similarly, AP/2/3/9 is dated 10 August 1829 whilst AP/2/3/10 is dated 6 August 1829. There are similar inconsistencies in book four of volume three and in volume four book one six letters from Arnold to various people are not arranged chronologically at all.

Book two of volume four seems to contradict the theory that the documents are arranged by the groups in which they were donated to the school. This book contains some 15 letters, all of them photocopies. The first is from Arnold to Blackstone, the original being in the National Library of Scotland. The following group of four letters, from Arnold to the Earl of Denbigh between September 1829 and October 1831, are copies of originals which are the property of the present Earl of Denbigh. The next group of three letters are from Arnold to John Gilson written between December 1825 and October 1826; again the originals are in the National Library of Scotland. The next two letters, from Arnold to H. Strickland, have their originals in the Gloucestershire County Record Office. The book concludes with two letters from Arnold to Reverend Hawkins and three to Doctor Hampden, all of which are copies of originals held in the library of Oriel College, Oxford. Although the individual letters within the sender/recipient groups are arranged chronologically there seems little likelihood that all these documents came to the School at the same time or from the same donor.

Whatever the shortcomings of the current arrangement, it was decided at the earliest stage of this project that re-arrangement of these documents would not be part of its brief. The individual letters are mounted with tape to the books and would not easily be removed without possible damage to the documents, and although many of the documents are photocopies there are still many original early nineteenth century items too. The objective, then, was to make the information held within these papers retrievable in their current state of arrangement.

Had the intention been to produce a traditional paper-based inventory then the decision not to re-arrange the documents would not have been easily reconciled with this intention, but with a computerized database solution the arrangement could be handled within the machine at the time of searching. As Woolgar says in connection with his Wellington Papers database project at Southampton University:

"...the sorting and arrangement of collections have traditionally been regarded as a highly refined art. What case is there for sorting a collection out of the order in which it is found if the catalogue can be re-ordered instantaneously to suit each researcher by the researcher himself ? (1)."

3.2 Methodology

The most important element to bear in mind when it came to the practical planning of the methodology to be employed in creating a finding aid for the Arnold Papers was the client's wish that the aid should make use of the convenient data storage and manipulation facility provided by her existing laptop personal computer (PC).

Back in 1971 Kenneth Darwin, then the Deputy Keeper of Records of Northern Ireland, stated that the most important impact of computerization on the field of archives administration was in the production of the hardcopy print-out catalogue, by taking in data from disparate sources and synthesizing and sorting it into a comprehensible order (2). For example, names from many parish registers could be entered, from a retrieval point of view randomly, and integrated lists of names could be printed out in alphabetical order covering all the registers entered (3). Fundamentally, according to Darwin, we are

"...talking about using the computer to provide finding aids not to do individual searching in records (4)."

Darwin was obviously writing at a time before the advent of very powerful, affordable desktop computers. He was also responsible for the records of six counties, so had volume of data very much in mind and, consequently, mainframe computers incapable of real-time interactive dialogue with the researcher.

The great strength of the computer is its ability to process huge quantities of data at speeds which in a human frame of reference can be thought of as instantaneous. Darwin comes close to recognizing the revolutionary qualities of this processing power:

"We must not just put information into a computer *to do for us what we know can already be done manually*: like the creation of a substitute for the usual kind of card index. We must not put information into a computer in order to have it produced in three conventional ways; we must put it in and expect to be able to pull it out in 25, 2,500 or 25,000 different ways (5)." (Author's italics).

The important aspect of this passage is in its admission that to use computers to do what we can already do is, by implication, a waste of the potential processing power of the computer. Darwin seems not to have taken the extra step and questioned the accepted methods of description, but merely sought ways in which the computer could help archivists to do what they had traditionally done, only more efficiently. To use an analogy, it was as though Gutenberg had used his press to produce a handful of bibles. The printing press only makes sense when producing books in large quantities, otherwise it could not compete with hand written manuscripts. The computer only makes sense when employed in the way which uses its strengths; speed of processing and accuracy. As long as what you put in is accurate then what you get out will be accurate in whichever way it is retrieved. Computers do not make mistakes.

By 1988, when Woolgar was writing about the Wellington papers database project at the University of Southampton, the importance of the computer in information storage, manipulation and retrieval was more apparent:

"The objectives of the project were, first, to find a system capable of accepting descriptions of all the Wellington papers to form an electronic database, which would serve as the principal means of reference to the archive, available *for on-line interactive searching* (6)." (Author's italics).

So, whilst Woolgar questions (*perhaps slightly tongue-in-cheek*) why we should sort a collection from the order in which it is found, when the data can be sorted all but instantaneously by computer (7), Jonathan Pepler, of the Cheshire Record Office, argues that this is insidious as it implies that the documents are all individual random items: "which can be re-ordered in any convenient way without affecting the nature of the collection (8)."

As is often the case the most productive path would seem to lie somewhere between these two positions. Whilst a large public access archive patently needs a browsable surrogate for its archives a small repository such as the Rugby School library is in no such position. The librarian is the conduit through which the vast majority of enquiries are made of the archives. The need in this case is for information on specific topics to answer specific questions. This means that accurate retrieval of information in as flexible way as possible is the prime requirement of a finding aid. This has been provided by indexing at a very detailed level and building a database containing the garnered knowledge from the archives. When contemplating the computerization of archival finding aids, as with a computerized solution to any manual process, the question to ask is not 'how do I use the computer to mimic a manual system already perfected ?' but 'how do I use the capabilities of the computer to fulfil the specified task as efficiently as possible ?'.

In this case the task was not to consign a traditional paper-based finding aid to computer as an expensive and wasteful ring-binder. It was to use the data manipulation capabilities of the machine to provide an electronic representation of the information

contained within the archive, so that it might be sorted, searched, listed etc., in as many ways as possible. It also enables the updating of records when enquiries are made and answered, as well as being readily expandable to the rest of the Arnold Papers and to other collections within the archive.

The differences in approach to the problem of archival description can be represented diagrammatically to illustrate the points of departure between my approach and that of traditional, paper-based finding-aids:

<u>High level tasks</u>	<u>Traditional</u>	<u>Traditional + computer</u>	<u>My approach</u>
Preservation and access			
moral defence	arrangement	arrangement	analysis of
sanctity of record gp.	description	description	data and user
			needs
classified list	printed	prepare data for	
		input and output	store in
indexing	printed/bound lists,		databank
calendaring	indexes, calendars	use machine to	
finding aids	etc.	sort in conventional	many on-line
sectional lists etc.		way	search strat-
			egies with
ACCESS	ACCESS	Print-outs of	option to print
		consolidated lists	
			ACCESS
		ACCESS	

For this finding aid the hierarchical arrangement has largely been set aside. The hierarchy is there in the assignment of reference codes but I have not followed the course of providing description according to this model, with increasing level of detail at lower levels. The way in which a relational database works best is, as its name implies, when data are arranged within it in terms of their mutual horizontal relationships. This aspect of the methodology will be discussed in more detail in chapter five when the database design is dealt with.

The levels implicit in any archive are, however, captured in the reference coding system I have chosen. Along the lines outlined as an example of description in the previous chapter I have chosen the letters AP to identify the Arnold Papers. As the

finding aid is not earmarked for use outside of the archive itself a designation for 'Rugby School' is not needed. The arrangement which has been inherited by this current project is of books within volumes. Therefore, the next levels reflect this arrangement: AP/2/3 would identify volume two, book three. Within the individual books the pieces are numbered consecutively from one as they appear. So, AP/2/3/12 is the twelfth piece in book three of volume two.

These reference codes provide unique 'names' in the repository and the finding aid for each piece. Within the database the reference code provides the link between each of the five tables which make up the database.

In short, then, the aim of the methodology outlined above is to contradict the following comment of Pepler's made in 1990:

"There are few if any effective applications of relational databases in the field of archive listing at present (9)."

3.2.1 Data gathering

Although it is important to have a fairly fixed idea of what data one needs and how it is to be used before the gathering begins there is, almost inevitably, some degree of evolution in the plan as the data gathering takes place. One is in the position of needing to analyze the data before being able to settle on the way to describe the collection, but having to gather the appropriate data before a detailed analysis is possible.

The approach taken for this project, then, was to gather from the very beginning any data which might conceivably be needed. This approach soon developed into a more or less standardized way of noting the data from each piece. The data needed can be split into three kinds: that describing the physical attributes of the document, that describing the information content of the document and that fulfilling the records management tasks of retrieval and storage.

The physical attributes noted were: what type of document is it ? Is the piece original or is it a photocopy or typed copy ? How many pages make up the piece ? What

material is the piece made of ? What size is it, in centimetres horizontal measurement x vertical measurement ?

Most of these items of information on the physical make-up of the documents are needed to manage the collection, rather than for the retrieval of information. The type of document needs to be known so that the archivist and/or researcher can draw what conclusions they think necessary from the type of document as regards their enquiry. The size and number of pages can be used to keep track of the physical integrity of the documents, in terms of loss or theft of bits of individual pieces. Size is important if one is thinking of mounting or displaying items. Finally, the material of which the document is made can be of importance in researchers' evaluation of the document, and is an important consideration in terms of production, storage and conservation.

The items of data gathered for the information content of the collection were: for letters, which make up the overwhelming majority of the Arnold Papers, the sender, recipient and date of writing, not the date of postmark or receipt. Then, for detailed indexing purposes, all people, places and subjects mentioned in the text of the letter. People and places are quite straight forward, but subjects are more...subjective. For concepts which appear often, I tried to develop a standard way of noting this. For example, the Newmanites were also referred to as Tractarians and Apostolicals. . Searches can be carried out using either of the terms, 'Newmanites' or 'Tractarians', as the index term for this concept is structured as follows: Newmanites / Tractarians. So, either search would find the same information.

Where a letter is predominantly about a single theme, or such a theme is a substantial proportion of the letter's text, then an abstract or synopsis of the letter was noted. Many letters, however, ranged widely and from the archivist's point of view randomly over subjects and family gossip. In these cases no abstract was noted and the abstract field in the database was left blank.

3.2.2 Data analysis

The data gathered according to the guidelines outlined above needs to be analyzed. This means that the pattern of the finding aid to be produced will largely be a function of the data drawn from the collection. With the finding aid in this case being a computerized solution the need was to represent the information within the archives through the data gathered and its organization within the database. The database forms a model of the collection; a model which can be viewed from any angle the enquirer wishes through the medium of the computer's manipulation of the representative data. The use of a computer to re-arrange data within a database for archival finding aids is analogous to that of three dimensional rotating views in computer-aided design when compared with flat, printed blueprints. How the analysis of the data gathered went to influence the design of the database will become apparent when the process of database design is explained in chapter five.

3.2.3 Description

Chapter one of this dissertation included a section enumerating the objectives of the project. These were fivefold: to provide a useful and usable finding aid, computerized infrastructure for collection management, to realise the advantages of a computerized solution and to produce a solution with flexibility to fulfil all the roles of finding aid, collection management tool and infrastructure for expansion. These are the prime movers of the attempt to reconcile user needs with the principles of archival description discussed in chapter two.

The purpose of description is to allow for the exploitation of a collection. An important part of that purpose is served by the physical manifestation of the description which is really a browsable surrogate for the holdings themselves. The Arnold Papers form a homogenous group in terms of their origins, but there is little logic apparent in their creation. That is, they do not represent the formalized proceedings of a committee such as a series of minute books would. They do not follow logically on one from another as the administrative documentation of an organization would. The information is in the

form of nuggets within an aggregate of fairly structureless writing. The description, therefore, has been designed to draw out the information, rather than to provide an overview of the subsets of the collections for the enquirer to decide whether or not to consult the documents. Whilst the description undertaken is not meant to provide an alternative to consulting the documents themselves, it is designed to extract the information one needs in order to make an educated decision as to whether or not to resort to the documents. For example, an entry in a description such as this would give little intimation of what information the documents contain:

AP/2/3	Volume two, book three - letters	1803 - 1809
1 - 15	Letters from Thomas Arnold to his aunt Susan Delafield written from Warminster Grammar School, Wiltshire. Photocopies. Originals held in the library of Winchester College.	

The nature of letters is that they contain much gossip and unrelated communication. The need for a browsable hard-copy is not so strong at an archive such as that at Rugby School, as virtually all searching is done through the medium of the librarian herself. Enquiries are mostly received via the post or telephone and, as such, demand answers to specific queries. A researcher might be better equipped to judge the value of high level descriptions to topics which are central, or even peripheral, to their interests. This is not the case, however, with the archivist doing the searching.

This concern with the type of documents in the collection and the circumstances of the repository in question is of fundamental importance when describing archives. The client wanted a finding aid which would allow her to extract information (on where to find information !) whilst still on the telephone to the enquirer. It was also part of her specification that she should be able to update records with the results of any previous enquiries on a similar topic. This specification called for a flexible, quick, non-hardcopy description where specific detail was the product of the search.

The description of the Arnold Papers for Rugby School was not compromised by any necessity to fit in with a house style or to complement existing descriptions. This

project, and a similar one to describe the Vaughan Papers being undertaken in parallel with this one, has as one of its objectives the setting up of an infrastructure for description which the School can follow at leisure to expand the coverage of the holdings in its archive. The position of pioneer in this field carries both responsibility and freedom: responsibility to provide a usable infrastructure which can also be readily expanded, and freedom to do so without being constrained by existing systems.

The finding aid stemming from the above premises is arranged around a twin role. The first role is a description of each individual piece, from AP/1/1/1 to AP/5/5/7. This part of the finding aid, the table ITEMS in the database (see database design, chapter five), provides the information on the physical aspects of the collection, custodial history, provision for details of any previous enquiries, any restrictions on production and retrieval information.

Another table, LETTERS, is, essentially, an elaboration of the ITEMS table. It covers just the letters in the collection, as they account for the vast majority of the documents in the collection (233 letters from a total of 253 items). By having a separate table for the letters another point of access to the information in the database is provided. This table contains information on the sender and recipient and any abstract if there is one. The second role of the description is the detailed index. In the database this takes the form of three tables; one for SUBJECTS, one for PERSONS and one for PLACES.

The design of the description is intended to provide a wide variety of access points; through the index, through reference number or date of writing in the ITEMS table, or perhaps through the sequence of sender and recipient data held in the LETTERS table. The entry points are legion, and will be discussed more thoroughly in section 5.4 on using the finding aid.

3.2.4 Indexing

As mentioned above, the Arnold Papers finding aid is indexed on people, places and subjects. Persons are identified by last, middle (where known) and first name. A field is included for nickname because Thomas Arnold and his wife Mary had nicknames for all their children and these are used frequently in their letters. Another field added to this index because of the idiosyncrasies of the collection is one which identifies the person being indexed in terms of their relationship to Thomas Arnold.

Places are indexed by street/area/building, town/city, county and country; or any of these which are appropriate. The subject index is at once more simple and more complex. It is more simple because it merely contains 'subject'. There is, however, the problem of reconciling subjects or concepts developed at length within documents with one or two word entries in the subjects index. To help with this process of reconciliation I have combined points four and five of Storey's ten points to bear in mind about any kind of archive indexing (9), as listed in chapter two above. Each entry in the database index has its own record (point 4) but multiple records have also been used to cover two or more interpretations of a single subject (point 5). For example, the piece AP/5/1/5 talks about the Church, the State and the 'Church and State'. In the letter such divisions are not explicit, but in the index three entries have been made: Church, State, Church and State. This is to provide a reasonable number of entry points to this piece of information.

The subject index does differentiate between concepts and specifics. For instance, entries are made under 'Education', for instance, where the subject of education as an abstract concept is discussed, such as the pros and cons of science teaching in schools. When practicalities are discussed, such as the arrangements for the new London University, then entries are made under 'London University' etc. There is no 'doubling up' of index entries. If a searcher wished to include 'university' under the banner of education then a search of 'education' and/or 'university' would be appropriate.

Regarding the other points of Storey's checklist (10), the Arnold Papers have been indexed in as much detail as resources (time, manpower and money) would allow. The limited time available for completion of this dissertation has meant that only the first five

volumes of the Arnold Papers were included in the project so that the level of detailed indexing would not be compromised on later volumes through pressure of time. The second point on Storey's list is to establish indexing criteria at the beginning of the work. This was done. Necessarily one must know what items to note for indexing when working through the documents, or a second recourse to those documents would become necessary. The third point, of choosing a system with flexibility and room for expansion, is a fundamental objective of the project. To provide a computerized infrastructure for the continuation of the description of the Rugby School holdings was one of the five objectives outlined in chapter one above. The mere fact of representing the information held within the collection as data within a database means flexibility. Data can be sorted and manipulated before retrieval, increasing the number of access points to the information greatly.

Point six is simple enough; bring out as many proper names as possible. The PERSONS table contains some 1229 records and the PLACES table contains 893 records. With the proviso that inevitably some will have been missed, these figures will show that if a person or place appears in the indexed volumes it will appear in the index. These 2122 records of the people and places indexes come from a collection of 253 items, so, on average, some 8.4 persons and places have been indexed for each document. This is not including the 242 records in the SUBJECTS index.

Point seven of Storey's list is to distinguish between references to the list itself and to the documents. This is redundant in the case of the Arnold Papers, as there is no list. All references are to the documents themselves. The reference number is a coded version of where to find the piece whilst the 'where to find' field spells out in words where the document is held; e.g. volume 3 book 2, p. 13. Point eight sums up nicely the reason underlying the project. The memory of the Rugby School librarian, excellent though it is, is not upto the task of finding the relevant information for all the enquiries received; hence the need for a finding aid.

Point nine of Storey's list is fundamental to my interpretation of what a finding aid and what an index can be. Indexes can, and should, become works of reference in their

own right. That is precisely the philosophy underpinning this database. The indexes direct the searcher straight to the documents, not to a list or catalogue. The final point in Storey's list is that indexing should not be looked on as doing the work of the researcher for her, but as merely making it possible for her to do her own work. This system of computerized indexing is designed to use the facility for sorting and manipulation of data in order to provide the searcher with many access points to the data so that she may satisfy any enquiry she might have within the limitations of the collection itself.

References - Chapter Three

1. Woolgar, C.M. 'The Wellington papers database: an interim report'. *Journal of the Society of Archivists*. Vol. 9, no. 1, January 1988.
2. Darwin, Kenneth. 'The use of the computer in indexing records'. *Journal of the Society of Archivists*. Vol. 4, no. 3, April 1971, pp. 229 - 240.
3. *ibid.*
4. *ibid.*, p. 235.
5. *ibid.*, p. 238.
6. Woolgar, ref. 1.
7. *ibid.*
8. Pepler, Jonathan. 'The impact of computers on classification theory'. *Journal of the Society of Archivists*. Vol. 11, no. 1, January 1990, p. 31.
9. Storey, Richard A. 'Indexing archives'. *Indexer*. Vol. 5, no. 4, Autumn 1967, p. 168.
10. *ibid.*

Chapter Four - Computerization of archival description

4.1 Introduction

The PC software package used for this project is Microsoft Access version 1.0, a relational database product. This package was chosen by the Computing Master of Rugby School for purchase by the school to meet its computerized information management needs. This purchase had already been decided on before the current projects to produce finding aids for the Arnold and the Vaughan Papers were instigated. It was, therefore, all but unthinkable that the school would purchase a second package of a similar nature to Access for the archival finding aids. So, recommendation of software was not included as part of the brief for this project. The purchase of Access was brought forward so that it might be used for the finding aids.

This *fait accompli* was no great hardship for a number of reasons: firstly, a systematic review of the software available, both bespoke archival packages and commercially available database packages, would have been very time-consuming and consequently, contradictory to the aims of providing a usable finding aid and an infrastructure for future expansion. As it was, the need to implement a system from scratch meant that the volume of archival material included within the finding aid was compromised in order to have sufficient time to provide the infrastructure. There really was insufficient time to preface the construction of the finding aid with a thorough enough evaluation of possible software as to recommend to the librarian of Rugby School the expenditure of several hundred pounds over and above the cost of the package already being purchased.

Secondly, it is more in keeping with the realities of the world that one has to work with the tools with which one is provided. By this I mean that in many repositories the package used for archival description, once chosen, is a given for any new computerization project. It is as well to be able to work within what is a quite reasonable framework as to be able to pick and choose one's tools. Finally, the package itself, Access, is an up-to-date relational database package with a very powerful and flexible

array of capabilities for information storage, manipulation and retrieval and, as such, eminently suitable for the task in hand.

4.2 Evaluation of the software

The software package used in this project is Microsoft Access, a relational database product. A database is, as its name implies, a collection of data; an electronic version of a filing cabinet or a card-index box-file. Access is, strictly speaking, a database management system (DBMS). It allows the information held within the database to be manipulated; to be sorted, alphabetized, categorized etc.. The *relational* part of the title refers to how the data being managed needs to be organized. In this case *relational* implies data arranged in tabular form:

"A relational database management system is one that is specifically

designed to manage information that's organized into one or more tables. (1)"

It is the breaking down of the raw data into groups of data readily organizable into tabular form which is the principal challenge of the database design. The separate tables into which the data is organized are then linked by the *relationship* between key fields within the separate tables. These tables may be linked on a one-to-one relationship, or on a one-to-many relationship, or a many-to-many relationship. This concept of relational linkage between tables will be explained fully in the following chapter when the design of the Arnold database is discussed. The advantage, however, springing from this arrangement is that search strategies may find information held within disparate tables, linked by the related field. So, for instance, in the Arnold database (see fig. 1) the field **reference number** appears in all five tables which constitute the database. In the tables ITEMS there is only one occurrence of each **reference number**. In the related tables of SUBJECTS, PERSONS and PLACES the **reference number** for a particular document appears as many times as there are index entries from that document. So, if a particular

letter contained references to seven people who were to be included in the index, then there would be seven entries of the **reference number** for that letter.

The linkage between similar fields in different tables, in this case **reference number**, is, in the example, a one-to-many relationship. The reference appears once only in the ITEMS table, but may appear many times in the SUBJECTS, PERSONS or PLACES tables. A search on one of the three index tables may throw up one or more references. With the link-to-the ITEMS table these references, using the **reference number**, may be cross-referenced with entries in the ITEMS table; entries which only have to appear the once, rather than many times were they to be included within the index tables. So, for instance, references identified in the index tables can also find the **date written** field in the ITEMS table and link across to the LETTERS table to find the sender and recipient information and/or any **abstract** for the letter. This is just an example of the type of flexibility a relational database allows in searching for information. Many other search strategies are possible across the tables in the database.

Querying the tables is facilitated by the provision of a query system which creates dynasets (2). Dynasets are dynamically allocated files of information forming a subset of the database contents which fulfil the criteria of a query made of the database. Dynamically allocated means that, unlike with most other database packages, the query results are live data, not a mere copy of the data. This means that updates may be made to the query results and become effective immediately.

Dynasets can be used mini databases for further queries to narrow down the search and gather more specific information. Access supports Boolean logic searching, with the use of AND and OR operators to make searches more accurate and specific. Queries may be created afresh on-line, or they may be saved as pre-programmed searches. So, if there are frequently used searches they may be refined and saved, and a report may be written, so that the output is formatted and presented in a way chosen by the enquirer. Results may be shown on-screen or printed off for hardcopy reports.

Similar to the formatted output of reports are formatted data input screens known as forms. Forms allow screens to be set up so that data may be input to the database in a

user-friendly way, one record at a time. For an example of a form see figure 11 in chapter five.

References - Chapter Four

1. Simpson, Alan. *Understanding Microsoft Access*. Sybex, Alameda CA., 1993, p. 7.
2. *ibid.*

Chapter Five - Development of the finding aid

5.1 Introduction

In planning any computerized solution to a task previously carried out manually one needs first to examine the requirements the client has for the manual mechanism already in use. What is the problem which needs solving and what information flows are there which can be profitably computerized ?

In the case of the Arnold Papers the information flow is very simple. Simplicity would normally imply an optimum solution which should not be changed. In this case, however, this does not follow as there is no manual solution as such. The need is to locate information within the Arnold Papers which will answer enquiries received. There is no system available to the librarian at Rugby for doing that aside from the vagaries of memory and trial and error.

The need for the finding aid, then, is not to replace an existing manual system, but to provide a system in the first place. The data flow for such a system would be:

1. Request received by post or telephone.
2. Consult database interactively with search criteria.
3. Receive report from database (on-screen or as hardcopy).
4. Use report from database to find and consult appropriate document(s) in the Arnold Papers.
5. Information from Arnold Papers relevant to enquiry.
6. Answer enquirer via telephone or post.
7. Update **previous enquiry** filed of appropriate record(s) of ITEMS table.

These seven points constitute the fundamental information flows of the simple system envisaged to provide a finding aid for the Arnold Papers. The only paper-based information in this flow are the original requests and the written answers.

Knowing the information flow within any system is the key to providing a computerized solution to that system. If it can handle all the information flows then the solution is, fundamentally, sound. Naturally, a great deal of the functionality of the solution will come from the detail which fleshes out the bare bones of the information flow list and the design of the database itself.

5.2.1 Functional specification

As intimated above, the principal problem to be solved by the provision of this finding aid is the answering of specific enquiries with specific answers. In essence this means finding the relevant information, if it exists within the collection, retrieving the appropriate documents, answering the enquiry, updating the database with details of the query and how it was answered and replacing the documents accurately in their volumes on the shelves.

As well as being able to locate information there is a need to be able to manage the collection itself. For this information must be available regarding the collection; information about the information if you like. This means data on the physical attributes of the documents and information pertinent to the collection as a whole or subsections of the whole. For example, whether particular items are originals or copies, any restrictions on production, what size they are and so on.

As the main function of the finding aid is to provide answers to specific questions then the need for a detailed index is more important than the provision of a facility for browsing. The information contained within the papers needs to be indexed on:

- People
- Places
- Subjects.

These indexes will form the principal points of entry to the information in the collection. The indexes need to point towards information on who wrote the document, when it was written, who it was written to, an abstract where one is appropriate, previous enquiries and any notes on the document itself.

From the records management point of view there is need of information to be held on what type of document the piece is (e.g. letter), where the document can be found, where the original is stored if the one in the collection is a copy, any restrictions on production and a custodial history of the item where known. There is also a need to include data on some physical aspects of the collection; is the item an original or a copy, what size is the piece, how many pages constitute the piece and what material is the piece made from ? All these attributes help with the planning of conservation, mounting for storage and/or display, and production of the documents.

5.2.2 Technical specification

To translate the functional specification outlined above into a working database design necessitates the preliminary step of creating a technical specification from which the detailed database design may be worked out. The technical specification lays down in precise terms what data elements will be needed to fulfil the criteria of the functional specification and what type of data each element may need to contain, such as integers, numbers to significant decimal places, alphanumeric characters only, maximum number of characters etc.. All this information needs to be known so that the correct field type may be allocated to the correct data element in the database, in terms of what data it can expect to hold and what function it will perform in the finished database.

In line with the types of information required, as outlined above, namely retrieval, records management and physical attributes information, I have laid out the fields to be included within the database under these three headings plus indexes. The parenthetical description relates to the data type within Access.

Retrieval info.

Reference number (text)
Abstract (memo)
Sender's honorific (text)
Sender's first name (text)
Sender's middle name (text)
Senders last name (text)
Sender's title/position (text)
Sender's address (text)
Sender's town/city (text)
Sender's county (text)
Sender's country (text)
Recipient's honorific (text)
Recipient's first name (text)
Recipient's middle name (text)
Recipient's last name (text)
Recipient's title/position (text)
Recipient's address (text)
Recipient's town/city (text)
Recipient's county (text)
Recipient's country (text)
Notes (memo)
Previous enquiries (memo)
Language (text)
Where to find (text)
Date written (text)

Records mgt info.

Type of document (text)
Restrictions (text)
Donated by (text)
Donated when (text)
Where is original ? (text)

Physical atts info

No of pages (number)
Format (text)
Size in CM (text)
Material (text)

Indexes**Subjects**

Subject (text)

Persons

Honorific (text)
First name (text)
Middle name (text)
Last name (text)

Nickname (text)

Title/position (text)

Relationship to Thomas Arnold (text)

Places

Street/area/building (text)

Town/city (text)

County (text)

Country (text)

The words in brackets after the field names are the data types. These describe the type of data which the field will take. By far the most common is text. The standard format for a text field is for between zero and 50 characters. If necessary the length of the field can be increased. Many of the fields for this database have been designated text fields because they need to contain non alphanumeric characters such as '?' when all or part of a field's data is unknown, also they need to contain characters such as '/', as, for instance, the **reference number** field does to accommodate items of the format AP/2/3/2.

The memo data type allows for a field to contain up to 32,000 characters. This data type has been used for the fields containing **notes**, **previous enquiries**, and **abstract**. These fields are ones which may require lengthy discourses to be entered, and can be treated as free-form text areas for paragraphs of text. The only other data type to be used is number. This data type is used when it is possible that calculations may be carried out using the data held in that field. Only numbers, including decimal points and minus signs, may be entered into this type of field. Although it is unlikely that calculations will be carried out on the **number of pages** field, it does provide some form of data validation during data entry (1).

Other data types available within Access, but not used for this database are: currency, counter, date/time, yes/no and OLE object. Currency, as its name implies, is used when dealing with data representing monetary values. Counter provides an incrementing number which relates to the number of records in a table, not to any

external reference. Date/time necessitates entry of data in one of a number of date and/or time formats, allowing chronological sorting to be carried out on that data. This data type was not used for the **date written** field in the Arnold Papers database because many of the items cannot be identified, in terms of date, precisely. For the facility to sort on the date/time field all the entries must be complete in the format chosen or defined. So, in the case of the Arnold Papers database the data dictates that the advantages of chronological sorting using the date/time data type must be foregone. The yes/no data type allows for a binary state field; on or off, true or false, yes or no. The OLE object data type allows for the importation of objects from other packages, such as scanned in images, for instance (2).

5.3 Database design

Having defined what data is to be included in the finding aid in the technical specification these fields must be organized in such a way as to hold their information in the tabular form necessary for entry into Access and in such a way as to minimize repetition of data but still link table with table and make retrieval of information from the database possible.

Figure 1 on the following page shows, diagrammatically, how the fields listed in the technical specification were divided into tables and linked together by virtue of common fields.

The Arnold Papers database is built around the central table, ITEMS. This table contains some 253 records; one record for every document in the first five volumes of the collection. Data included in this table are those which need only be represented once for each document.

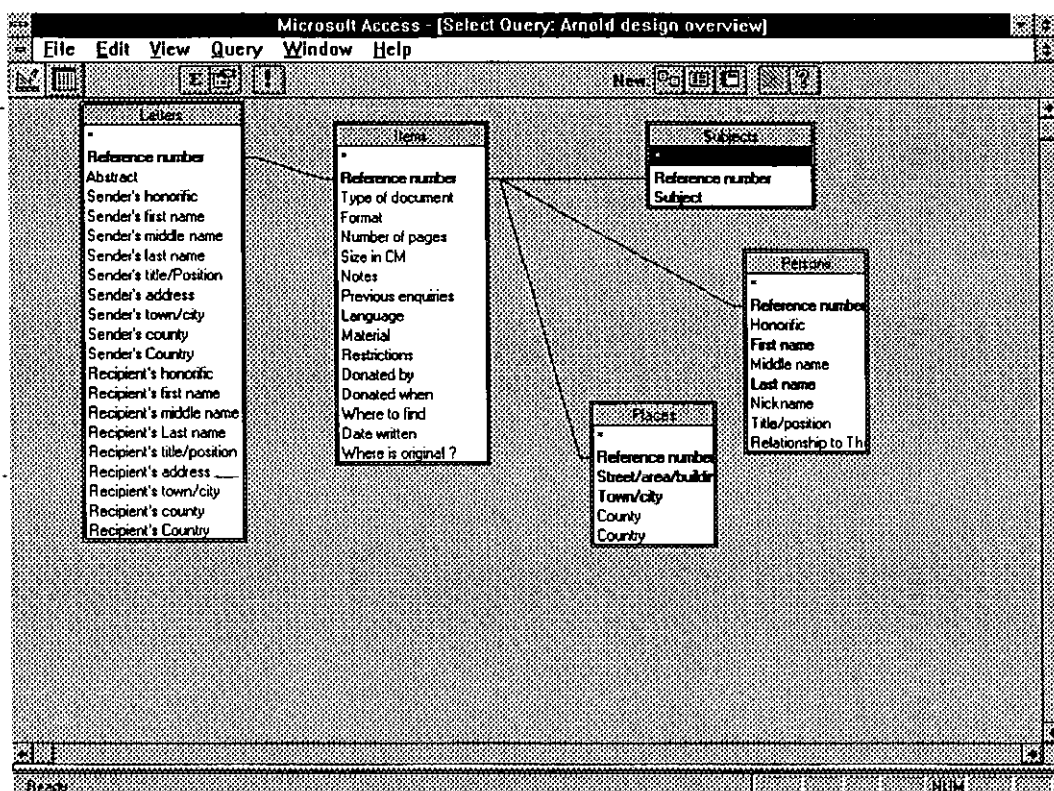


figure 1 - database design

The centre pillar of the database, the ITEMS table, is connected to the LETTERS table in a one-to-one relationship via the **reference number** field. As the vast majority of documents in the collection, 233 out of 253, are letters the provision of a separate table to deal with the specific peculiarities of letters suggested itself. As the letters are a subset of the items the relationship is clearly one-to-one. So, in LETTERS an item is described in terms of the details of the sender and recipient and, in some cases, an abstract of the letter's contents. The number of fields is necessary as the elements of name and address have to be broken down into their constituent parts to be able to use free text searching to find entries. This division of elements of names and addresses also facilitates sorting and listing by name, by town, by county etc. in alphabetic order. Any of these fields may be extracted in queries and arranged in any manner the enquirer sees fit.

The fields in the diagram which are in bold type are the key fields. These are defined in Access to provide each and every record with a unique identifier, or key, within that table. In the tables where **reference number** is at the one end of a one-to-one or a one-to-many relationship then it serves as the unique key field itself. Where it is at the many end of a one-to-many relationship then it has to be supported by other fields which will identify uniquely each and every record.

These key fields help speed up the retrieval of information by Access, as well as being necessary to the mechanics of the product. The reference field of the ITEMS table links, on a one-to-many relationship, with tables for indexing SUBJECTS, PERSONS and PLACES. So, for every **reference number** in the ITEMS table there may be any number in any of the three index tables, according to how many index entries there are for the document identified by that **reference number**.

The facility for browsing the collection's contents by means of a surrogate, as discussed in chapter two, has been incorporated within the design of the Arnold Papers finding aid. The tables ITEMS and LETTERS, in which each separate record in the table equates to a distinct document on the shelves, provide means of browsing the collection's holdings. The tables may be viewed in the design view in their entirety or, more helpfully, via a simple query giving the required fields. For example, the recipients' name fields selected in a query with no added selection criteria would enable all the recipients of letters to be browsed. It would also be possible to sort them by name in alphabetical order. By isolating the **date written** field of the ITEMS table the chronological arrangement of material could be browsed.

The browsing facilities provided by the one record to one document tables is supplemented by the ability to narrow the selection of data to be browsed by a selective query first, or, having found something of interest whilst browsing, a selective query could follow up by providing the **reference number** and any other chosen information on the selected records.

5.4 Using the finding aid

The most basic form of interaction between searcher and finding aid is via a straight forward query on a single chosen table. The easiest way of selecting a new query is to click on the new query icon on the menu bar (see fig. 2) once the database has been opened.

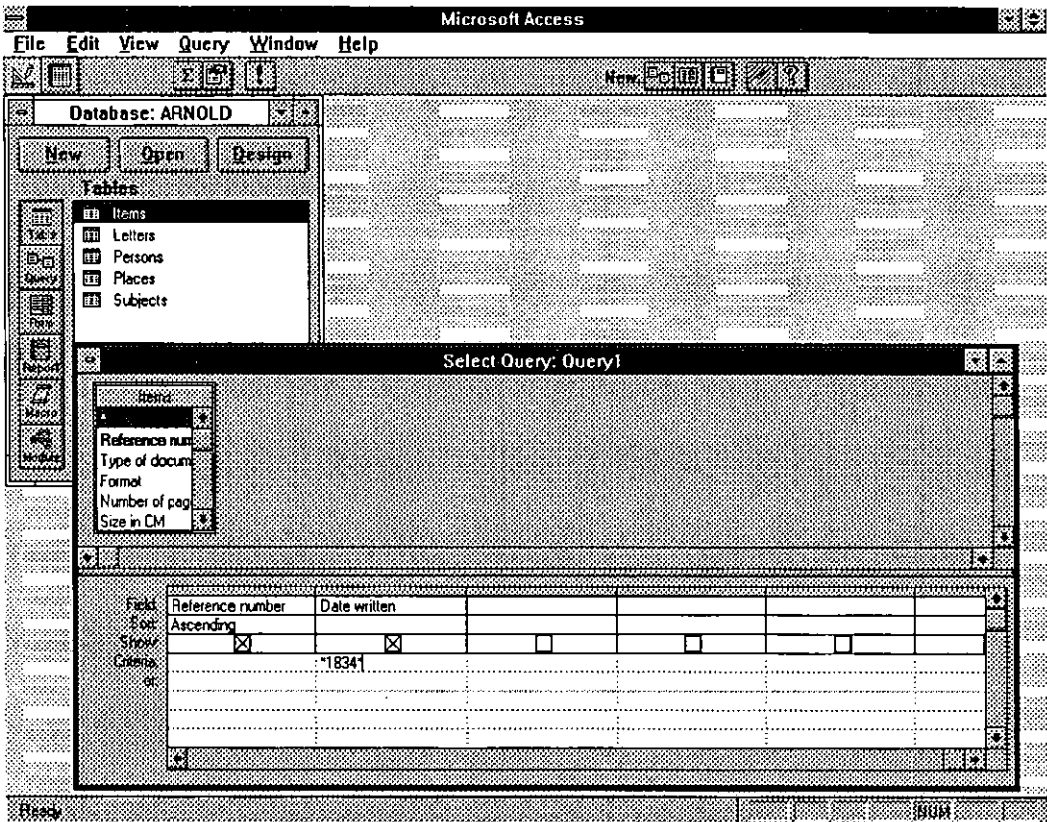


figure 2 - query example 1

In the above figure 2 the new query icon can be seen just below and to the right of the word 'Access'. The word 'new' is followed by a group of three icons and a group of two. The new query icon is the first one in the group of three. The table ITEMS is highlighted on the open database screen. Selecting a new query via the new query icon whilst a table is highlighted will select that table as the one upon which the query will be made. The select query window which is opened by selecting a new query includes, in this case, the fields from the ITEMS table as this is the highlighted table in the database screen.

For this example query I have selected the **reference number** and **date written** fields. This will produce a dynaset consisting only of records from these two fields from the ITEMS table. The **reference number** is there to identify which documents fulfil the criteria of our query. The query itself is on the **date written** field. In the criteria cell of the query screen under **date written** I have entered *1834*. The preceding and succeeding asterisks denote the query as looking for fields which contain the text 1834. If the asterisks were omitted the search would only throw up those occurrences where 1834 was the sole text within a field for a particular record. As the *1834* is in the criteria cell this identifies the query as being on that field. The **reference number** field in the dynaset will only contain the reference numbers corresponding to the records selected by the **date written** query. The **reference number** field is sorted in ascending order so that the records answering the query will be arranged by the reference number in order from the nearest to AP/1/1/1 up to the nearest to AP/5/5/7. Figure 3 below shows the result of the query in figure 2.

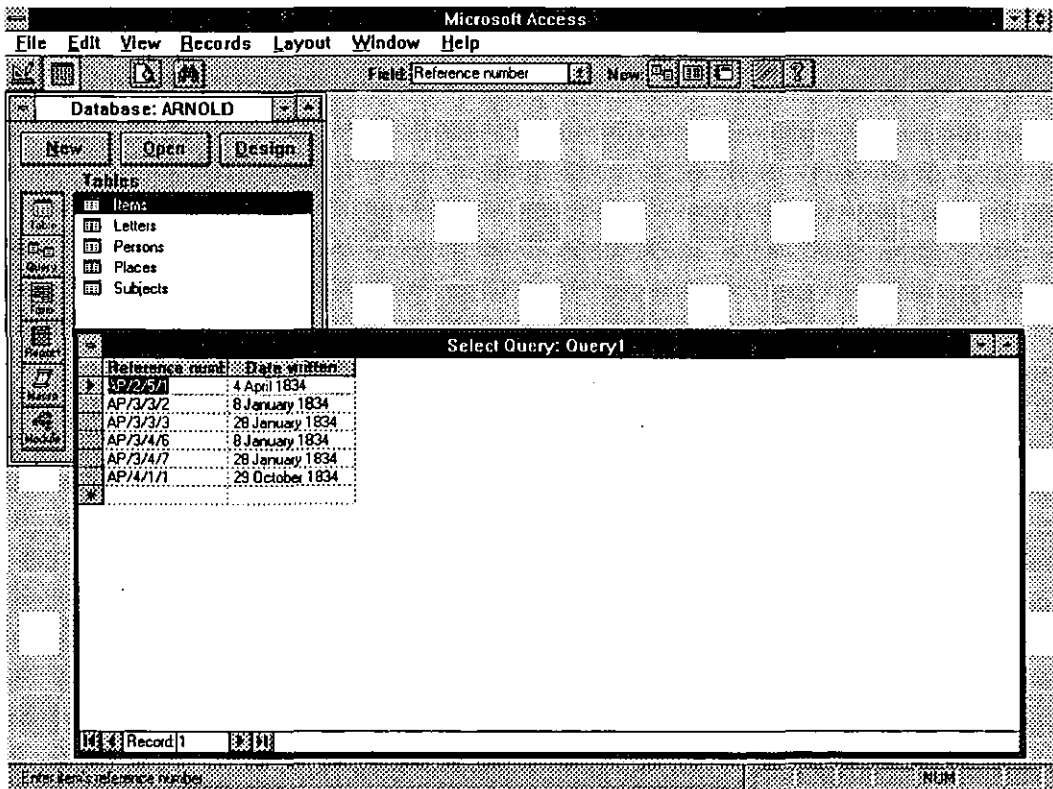


figure 3 - result of query example 1

Reference number	Date written
AP/1/2/11	[180?]
AP/1/2/12	[?]
AP/1/2/13	[?]
AP/1/2/14	[?]
AP/1/2/18	[?]
AP/1/2/21	[?] 1857
AP/1/2/9	8 November [180?]
AP/1/3/1	[?] 1803
AP/1/3/11	9 September [1806?]
AP/1/3/3	18 April [1804?]
AP/1/4/2	
AP/1/4/4	
AP/1/5/1	[21 August ?] 1819
AP/2/1/2	19 April 18 [?]
AP/2/5/1	4 April 1834
AP/3/1/14	[1831 ?]
AP/3/1/17	[July ? 1835 ?]
AP/3/1/20	29 June [1840 ?]
AP/3/1/9	8 May [1829 ?]
AP/3/3/2	8 January 1834
AP/3/3/3	28 January 1834
AP/3/3/4	[1841 ?]
AP/3/4/12	[December 1840 - Janua
AP/3/4/6	8 January 1834
AP/3/4/7	28 January 1834
AP/4/1/1	29 October 1834
AP/4/1/13	15 December 18[?]
AP/4/1/14	[?] 1835
AP/4/2/10	7 August [1827 ?]
AP/4/5/4	[?]
AP/4/5/6	[1843 ?]
AP/5/1/12	[3 ?] February [1828 ?]
AP/5/1/13	[2 ?] February [1828 ?]

figure 5 - result of query example 2

The query asking for records including 1834 or null value (blank) or characters within square brackets finds 43 records meeting these criteria. From those the ones which are plainly from a year other than 1834 can be rejected, such as AP/1/3/11 9 September [1806?]. Others, such as AP/4/1/13 15 December 18[?] might be investigated further to see if they come within the ambit of the search. There are no records which show a blank because for the **date written** record any unknown date was denoted with a question mark within square brackets. This is not the case with all records for which there is no entry, many are left blank. This is because **date written** is a field which has obvious merits for searching on, as well as being known in the majority of cases. Many fields, though, such as middle names are very rarely known and so mostly left blank. The country field in the PLACES table is only used when a place being indexed is outside England or not a well known British town. The null value search, then, is still a valuable tool to be borne in mind when instigating new queries.

In figure 4 the method of carrying out searches using the **or** operator can be seen. Under the cell marked 'criteria' is another marked 'or' and below this other unmarked cells. To make an **or** query the various criteria should occupy separate cells beneath the criteria cell in the column denoting the field the searches are being made upon; in this case the **date written** field.

To carry out an **and** query on a single record one merely links the search criteria in the same cell. So, instead of putting each separate criterion in the cell below as for **or** searching, **and** searches would be of the kind: *1834* and *[?]* in the same cell. This particular query would not be all that useful as it would only find occurrences of 1834 within square brackets. **And** operator searches are much more useful when linking separate fields, for instance when searching for *Susan* in the **recipient's first name** field and *Delafield* in the **recipient's last name** field of the LETTERS table to find letters addressed to Susan Delafield.

And and **or** operators may be combined across fields. For instance, to continue with the Susan Delafield example, the **recipient's first name** field might include the null value and/or *[?]* queries under the **or** cells to include any records where the first name is unknown or in doubt. As can be seen from figure 6, the combining of **and** and **or** operators across fields within a single table requires the repetition of the *Delafield* criterion in the **or** cells below the **recipient's last name** field. This is because **and** searches are so defined by placing the criteria in cells on the same horizontal level, whereas **or** searches are denoted by the criteria appearing in different vertical cells.

Were the 'Like delafield' entries not duplicated in three adjacent vertical cells the query would not work. This query is, effectively, asking three separate questions: firstly finding all the occurrences of 'Susan' and 'Delafield' (N.B. capitalization is not significant in the searching process), secondly all the null values under **recipient's first name** and 'Delafield' in the **recipient's last name** field, and finally characters within square brackets in the **recipient's first name** field and, again, 'delafield' in the **recipient's last name** field. As all the criteria rows are linked by the **or** operator the resultant dynaset gives the records which fulfil the characteristics of all three of the questions.

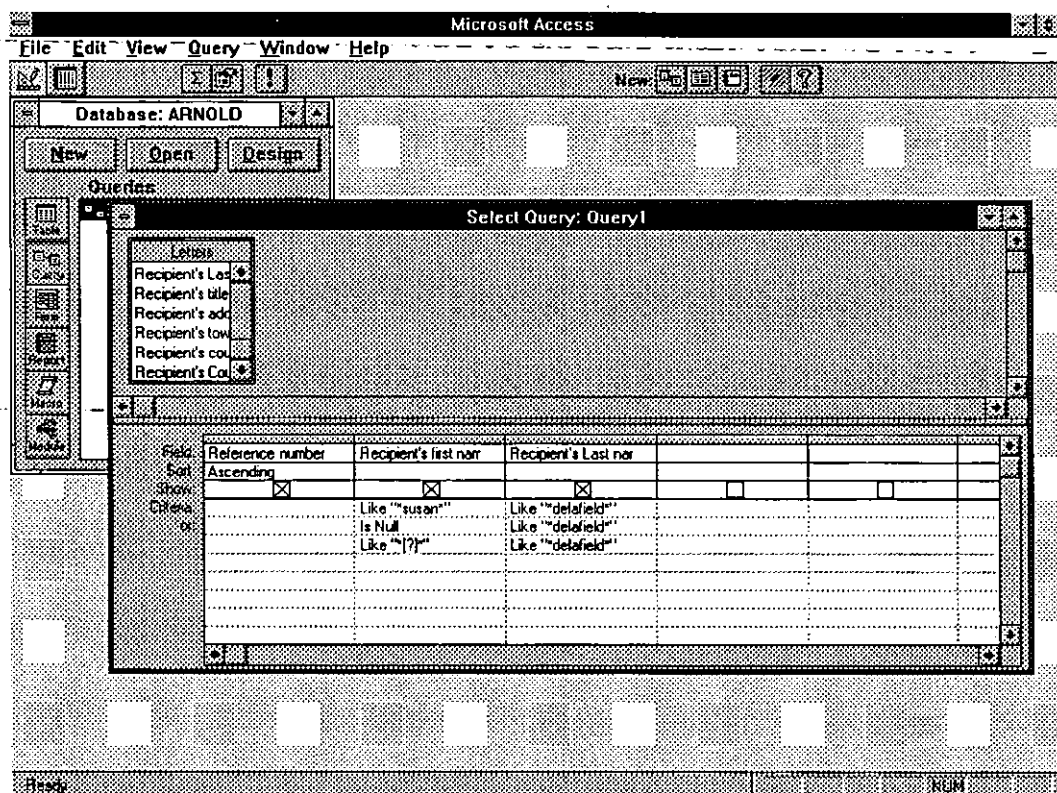


figure 6 - query example 3

When combining **and** and **or** operators one must think carefully about the questions being posed to the database. The grid of the select query screen can be thought of as asking different questions; vertically arranged criteria are utilizing the **or** operator and horizontally arranged criteria are utilizing the **and** operator (3).

The Boolean logic of **and/or** searching does not always follow the grammatical logic of the question posed in English (4). For instance, a question might be: "I want to find all the items written in the years 1828 and 1829." A search linking 1828 with 1829 in the criteria cell of the **date written** field in the **ITEMS** table would produce an empty dynaset because no item could be written in both 1828 and 1829. This question would have to be arranged as an **or** query. This is quite a straight forward example, but more ambiguous examples will inevitably show themselves when using the finding aid in earnest.

Reference number	Recipient's first	Recipient's last
AP/1/3/7	Susan	Delafield
AP/1/3/8	Susan	Delafield
AP/1/3/9	Susan	Delafield
AP/2/1/10	Susan	Delafield
AP/2/1/11	Susan	Delafield
AP/2/1/12	Susan	Delafield
AP/2/1/13	Susan	Delafield
AP/2/1/14	Susan	Delafield
AP/2/1/15	Susan	Delafield
AP/2/1/16	Susan	Delafield
AP/2/1/17	Susan	Delafield
AP/2/1/18	Susan	Delafield
AP/2/1/19	Susan	Delafield
AP/2/1/20	Susan	Delafield
AP/2/1/21	Susan	Delafield
AP/2/1/22	Susan	Delafield
AP/2/1/23	Susan	Delafield
AP/2/1/3	Susan	Delafield
AP/2/1/4	Susan	Delafield
AP/2/1/5	Susan	Delafield
AP/2/1/6	Susan	Delafield
AP/2/1/7	Susan	Delafield
AP/2/1/8	Susan	Delafield
AP/2/1/9	Susan	Delafield
AP/2/3/15	Susan	Delafield
AP/2/3/16	Susan	Delafield
AP/2/3/17	Susan	Delafield
AP/2/3/18	Susan	Delafield
AP/2/3/19	Susan	Delafield
AP/2/3/20	Susan	Delafield
AP/2/3/21	Susan	Delafield

figure 7 - result of query example 3

As can be seen above, in figure 7, the search shown in figure 6 has found 44 occurrences of Susan Delafield as the recipient of letters held in the Arnold Papers.

Another type of search which might prove useful, particularly as a way of making the index tables (PLACES, SUBJECTS, and PERSONS) more browsable is a type of search known as a unique values search. A unique values search differs from a normal search in that the latter normally displays all the records which meet the criteria of the search. The unique values search, however, only displays one of what would otherwise be duplicate entries (5). This type of search could usefully be employed in producing a browsable version of the PERSONS table which, in its entirety, contains some 1229 records. The resultant dynaset contains some 355 separate entries, meaning that the PERSONS table contains entries on 355 different surnames. This number of entries, arranged in alphabetical order are easily browsable for a particular name, knowing that if

the name is not in the dynaset then it is not in the index. This browsable unique values dynaset should be used as would an index to a list or catalogue; to look up a name and, if it exists, to follow up with a more specific and 'normal' query.

To set up a unique values query the field one wants to consult, in this case **last name** in the PERSONS table, must be the only field selected. Click on **View - Query Properties** and then click on 'Unique Values Only' in the query properties dialogue box. Then run the query in the usual way. A query may be run by choosing **Query - Run** from the menu bar or by clicking on the datasheet view icon in the top left-hand corner of the menu bar (under **File** and just to the right of the design view icon which looks like a set square and a pen).

The unique values search on **last name** in PERSONS has been saved as: 'Surnames included in the DB'. When using the unique values search method on new queries it is as well if, when the query is finished, the 'Unique Values Only' box in the query properties dialogue box is toggled off, as it remains in operation if another search is requested using the same grid without closing the query.

In a similar way to querying single tables multiple tables may be queried together. This is an important function of a relational database because connected data can be safely put in different tables if more appropriate without restricting access to it. For instance, the **date written** field was located by the database design in the ITEMS table rather than in the three index tables with the indexing terms. This was because the **date written** field only needed to be shown once for each item, whereas there are multiple indexing terms for each item. To have put the **date written** field with the indexing terms would have been a great waste of space and effort.

Figure 8 shows an example of a query on fields from multiple tables within the Arnold Papers database.

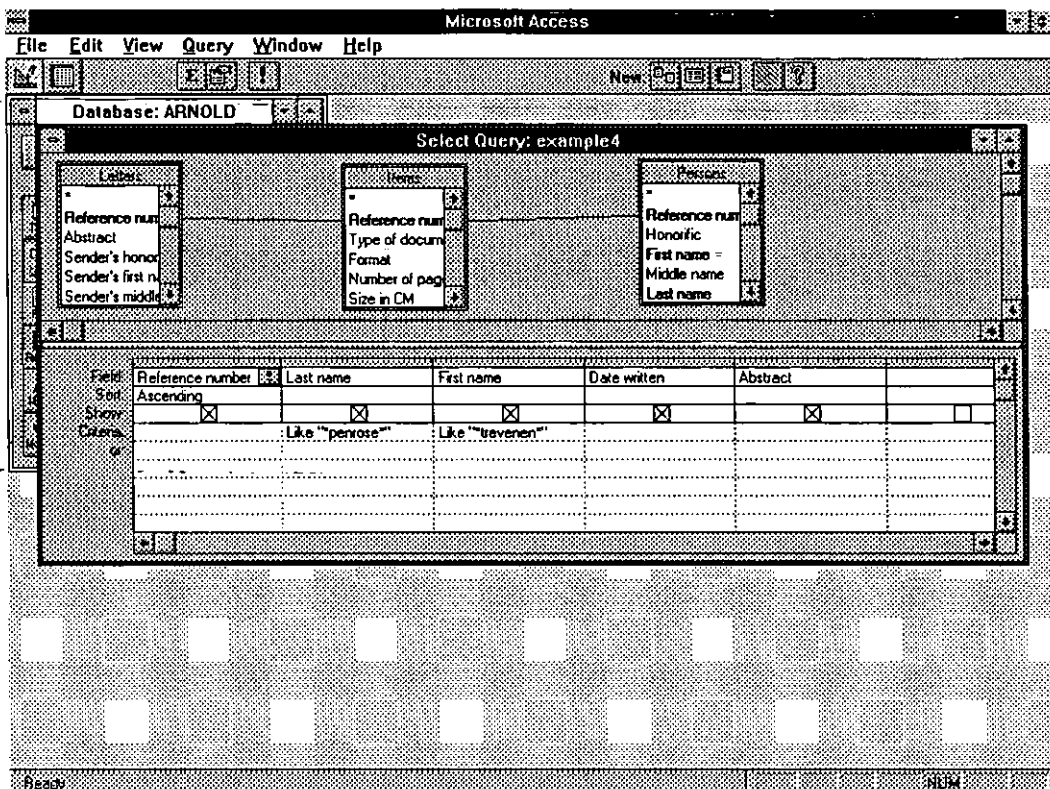


figure 8 - query example 4

To set up a query across multiple tables the individual tables concerned must be linked with one another via similar fields. This can be done at the time of the query or when the database is set up. It is easier and safer, in terms of making accurate searches, if the linkages are defined at the time the database is designed. This has been done for the Arnold Papers database, with all five tables being linked via the field **reference number** which appears in all five tables (see figure 1 database design for table relationship definitions). The new query is opened in the same way as a new query on a single table, click the query icon on the database screen and then click on the 'new' icon on the same screen. The 'Add Table' dialogue box then gives the searcher the opportunity of selecting the tables to be included in the query by highlighting them and clicking on the 'Add' icon one at a time. When all the tables required for the query have been added, then click on the 'Close' icon to return to the 'select query' screen.

As can be seen in figure 8, the tables selected are shown with the related fields linked by black lines. In the example query I have chosen the fields **reference number** and **date written** from the **ITEMS** table, the fields **last name** and **first name** from the **PERSONS** table and the field **abstract** from the **LETTERS** table. The resultant dynaset will be sorted by **reference number** in ascending order. The search terms: ***penrose*** and ***trevenen*** have been entered into the criteria cells for **last name** and **first name** respectively. Note that the software automatically translates ***penrose*** to: **Like "penrose"** and ***trevenen*** to: **Like "trevenen"**.

This query will search on **first name** and **last name** of the **PERSONS** table and use the link via **reference number** to display the **reference number**, **date written** and **abstract** fields for the records meeting the search criteria.

Reference number	Last name	First name	Date written	Abstract
AP/1/2/17	Penrose	Trevenen	20 September 1819	Thomas Arnold's fu
AP/2/1/10	Penrose	Trevenen	19 January 1813	
AP/2/1/12	Penrose	Trevenen	10 July 1813	
AP/2/1/13	Penrose	Trevenen	23 November 1813	
AP/2/1/14	Penrose	Trevenen	10 December 1813	Tells of Dyson's pla
AP/2/1/15	Penrose	Trevenen	11 January 1814	
AP/2/1/16	Penrose	Trevenen	14 February 1814	
AP/2/1/17	Penrose	Trevenen	30 April 1814	
AP/2/1/18	Penrose	Trevenen	24 May 1814	
AP/2/1/19	Penrose	Trevenen	22 June 1814	Description of the c
AP/2/1/21	Penrose	Trevenen	12 November 1817	
AP/2/1/23	Penrose	Trevenen	27 July 1820	Re: choice of cook
AP/2/1/4	Penrose	Trevenen	9 June 1812	
AP/2/1/9	Penrose	Trevenen	22 December 1812	
AP/2/3/10	Penrose	Trevenen	6 August 1826	
AP/2/3/12	Penrose	Trevenen	13 September 1829	
AP/2/3/2	Penrose	Trevenen	29 December 1823	
AP/2/3/20	Penrose	Trevenen	13 February 1812	
AP/2/3/3	Penrose	Trevenen	3 January 1824	
AP/2/5/2	Penrose	Trevenen	21 September 1838	The Newmanites. I
AP/2/5/4	Penrose	Trevenen	11 May 1841	
AP/3/1/12	Penrose	Trevenen	12 September 1829	
AP/3/1/18	Penrose	Trevenen	27 July 1836	
AP/3/1/6	Penrose	Trevenen	4 August 1820	
AP/3/1/8	Penrose	Trevenen	17 August 1827	
AP/3/3/4	Penrose	Trevenen	[1841 ?]	
AP/4/1/11	Penrose	Trevenen	31 March 1815	On Oriel election of
AP/4/1/12	Penrose	Trevenen	17 May 1830	Letter to Trevenen
AP/4/2/9	Penrose	[Trevenen ?]	4 June 1827	On Arnold's shortc
AP/4/4/4	Penrose	[Trevenen ?]	3 October 1838	

figure 9 - result of query example 4

The result of this query, shown in figure 9 above, is a dynaset containing all the references in the Arnold Papers database to Trevenen Penrose with the reference numbers of the individual items to identify the relevant documents, the date the documents were written and any **abstract** for the documents.

In the example query shown in figure 8 the three selected tables are linked, from left to right, in a one-to-one relationship between LETTERS and ITEMS and a one-to-many relationship between ITEMS and PERSONS. If one wished to query two tables which could only be linked by a many-to-many relationship, as would be the case with PERSONS and PLACES, for example, then a table must be interposed between them in order to bridge the gap between the two tables with duplicate entries in the **reference number** field (6). As can be seen from the database design overview in figure 1, the tables PERSONS and PLACES can be linked via ITEMS. So, for a query of the two former tables ITEMS would have to be included, though there is no need to view data from the bridging table if it is not required by the search.

The dynaset, or output created by a query, can be used as a subset of the main database in subsequent queries, meaning that a dynaset may be used as the data upon which a follow-up query is made. To do this the query supplying the dynaset must be saved. In the same way that the saved tables can be selected as the basis for queries so can saved queries be selected. Fields from the selected query can be chosen for the field cell to be queried. Sorting and search criteria can be added as for normal queries.

In figure 10 I have used the query I saved from the unique values search, 'Surnames included in the DB', to illustrate a follow-up query. The saved query, 'Surnames included in the DB' gives us all non-duplicate names held in the PERSONS table. Searching on **last name**, the only field used in the original query, the new query searches for *Arnold*, *Penrose* and *DeLafield* within the unique values subset.

The value of this specific search example is limited as the unique values output, in alphabetical order as it is, can easily be scanned for such information. It does, however, illustrate clearly, on a relatively straight forward original query, the facility to search on the output data of previous searches.

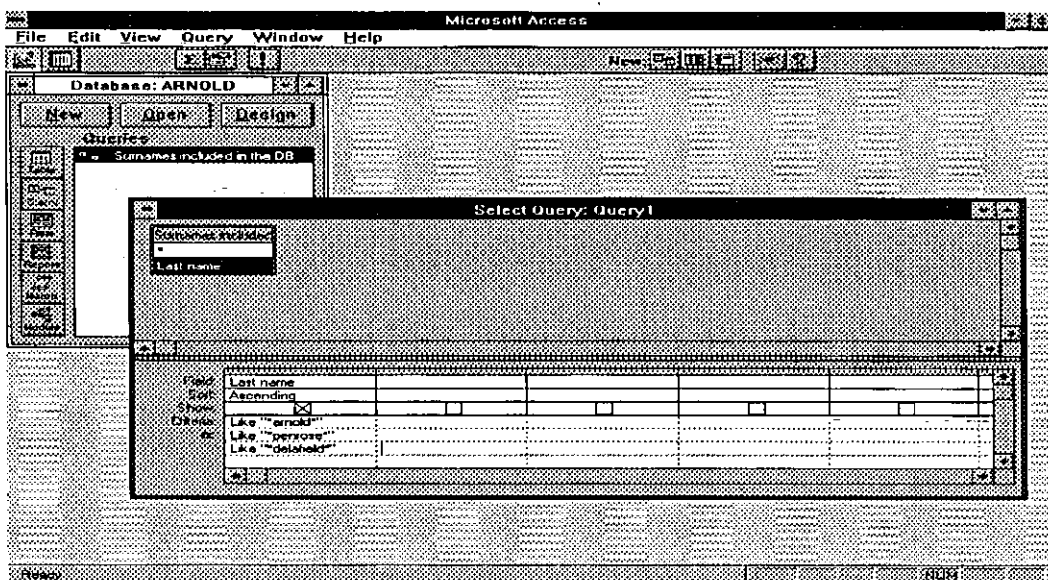


figure 10 - query example 5

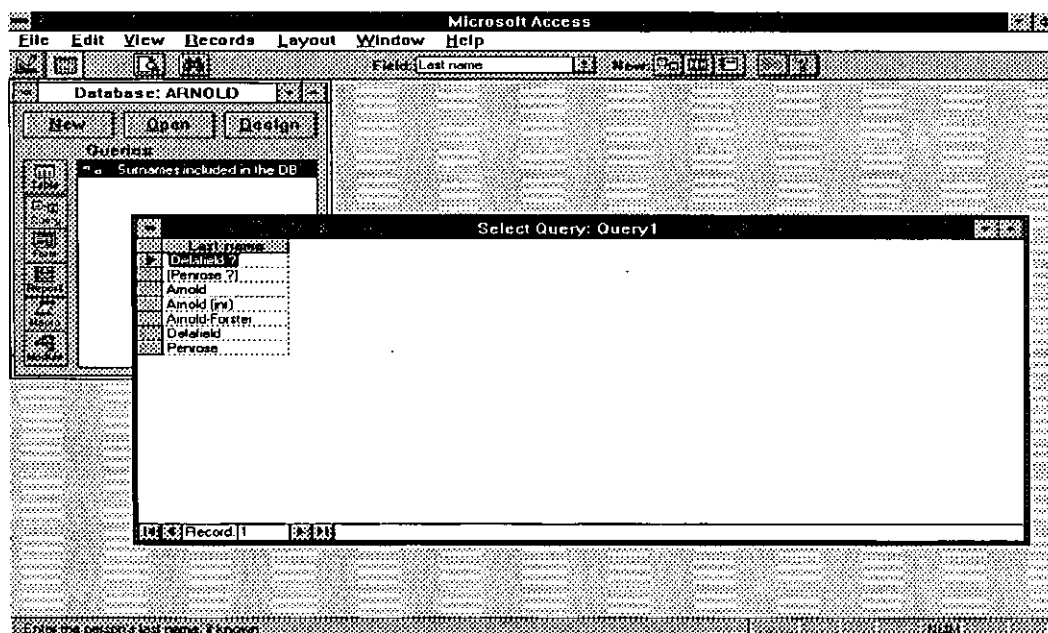


figure 11 - result of query example 5

As can be seen from figure 11 above, the dynaset resulting from the follow-up search has thrown up some variations on the search criteria which might have been missed by a visual scan of the unique values search dynaset. As well as the three names defined in the search criteria it has also found : [Delafeld ?], [Penrose ?], Arnold [jnr], and Arnold-Forster.

Microsoft Access

File Edit View Records Window Help

Database: ARNOLD

ITEMS data entry screen

ITEMS data entry screen

Reference number: LP/5/2A

Date written: 15 March 1839

Language: English

Medium: paper

Number of pages: 2

Size in cm: 18.5 x 23

Notes: Signature cut away

Provenance description:

Language: English

Medium: paper

Number of pages: none

Size in cm:

Notes:

Provenance description:

Record: 242

Enter item's reference number

figure 12 - ITEMS data entry screen

Figure 12, above, shows the form designed for the entry of data into the ITEMS table. Similar forms have been produced and saved, as part of the Arnold Papers database, for data entry to the other four tables.

These forms are for use in adding extra data to the appropriate tables. The screen above represents a single record in the ITEMS table, number 242 of 253. By tabbing from box to box one is prompted for the input data for each field and each record in turn. As a box is highlighted, in the example screen the **reference number** field is highlighted, a prompting message appears in the bar at the lower left-hand side of the screen to describe the type of data the form is expecting for the field currently highlighted. In the example the message is: "Enter item's reference number", appropriately enough.

5.5 Conclusion

The steps outlined in the previous section for the use of the finding aid: simple queries, selecting fields, sorting, search criteria, using operators and wildcards; **and/or** Boolean logic; viewing only unique values; querying multiple forms; carrying out queries on a dynaset of previous queries; and creating forms, are the fundamental skills of manipulating data within a Microsoft Access database. These skills provide a great pool of potential searching power and should suffice to meet almost any information retrieval problem one might encounter. All the methods dealt with in this chapter can be used in more sophisticated ways with experience, but this is not necessary to realize the utility of the finding aid.

The finished product, as described in this chapter above, provides a multitude of potential access points to the data stored within the database. Some of these I have tried to show with the examples chosen to illustrate usage of the finding aid. The ITEMS table contains a definitive listing of all the documents covered by this project. The linking of the five tables via the **reference number** field in each means that any piece of data in the database may be retrieved in terms of its relationship to any other piece of data. The three index tables, SUBJECTS, PLACES and PERSONS, can be used as entry points, and the linkage to the other tables allows data found within the indexes to throw up such information as: the date of the piece, where it is kept on the shelves, how big it is, what it's made from, what type of document it is, and so on almost indefinitely.

This ability to manipulate data in such ways is the great strength of the computer, and the flexibility it gives to searching for information represented within the Arnold Papers database fully vindicates the choice of a computerized solution for this project of providing a finding aid for the collection.

References - Chapter Five

1. Simpson, Alan. *Understanding Microsoft Access*. Sybex, Alameda CA., 1993, p. 88.
2. *ibid.*, p. 89.
3. *ibid.*, p. 264.
4. *ibid.*, p. 266.
5. *ibid.*, p. 271.
6. *ibid.*, p. 284.

Chapter Six - Conclusions

6.1 Problems encountered

The problems encountered in this project were of two kinds: those stemming from constraints of time, exacerbated by lack of experience and the fact that the project was beginning from scratch in terms of the Rugby School archives lack of existing provision for archival finding aids; and those stemming from the vagaries of the collection itself. These latter problems, it must be stressed, are indicative of a lack of experience on my part and not of a fundamental difference from the problems potentially posed by any archival collection.

The time constraint manifested itself in the need, a little way into the data gathering phase towards the beginning of the project, to adjust the scope of the exercise. It was originally intended to include the whole nine volumes of the Arnold Papers but, after gathering the requisite indexing data for the first five volumes a decision had to be taken to draw a line under the information gathering at this point. To have continued the data gathering to the remaining four volumes would have compromised the completion of the finding aid and necessitated indexing in decreasing detail for the later volumes. It was decided that it would be preferable to truncate the project scope and keep the same level of indexing detail, especially as a significant part of the project was the provision of a computerized infrastructure for the expansion of the archival description to other collections at Rugby.

The decision to index in great detail meant that every document in the collection had to be read through thoroughly. This in itself was very time consuming. The handwriting of the documents' authors, especially the almost illegible writing of Mary Arnold, was often difficult to read quickly, adding to the time needed for data gathering. The only answer to Mary Arnold's illegibility was slow deciphering of the text, luckily few of the letters in the collection were written by her.

This palaeographical problem of the writing links the two types of problem encountered. The public school way of addressing people, even close friends, by their

surname makes identification of people mentioned in the letters more difficult than it would otherwise be. This practice is exacerbated by the unavoidable phenomenon when dealing with letters written by strangers of having to interpret allusions which would be well known to the sender and recipient, so much so that there was no need to spell out the topic in the letter.

Other problems were caused by the arrangement of the papers themselves. These would not necessarily be found in other archival collections. A decision had to be made as to whether the existing arrangement should be kept or whether the collection should be re-arranged as part of the project. The mounting of the documents made their re-arrangement physically difficult and risky, whereas their logical arrangement was not as logical as it might have been. As they had been arranged, however, the original order of the documents had been lost. The evidence available for the custodial history of the various parts of the collection was piecemeal and haphazard. It was, therefore, decided to leave the arrangement as it was and proceed to use the electronic representation of the collection, the database, to 'arrange' the collection at the time of searching.

The unpredictable nature of an archival collection makes planning fields in a database to accommodate their data difficult. Database packages are designed for applications where all fields are knowable: telephone number, name, order number, customer number, date etc.. The problem with archives is that often pieces of information are missing, incomplete or, in some other way, unknowable. This means that fields have to be left blank or have indicators for missing or doubtful information. This constraint led to the use of the text data type in the database, allowing for any combination of characters to be entered. The biggest drawback in this was the loss of the facility to sort chronologically on a field defined as a date/time data type. As illustrated in the previous chapter, however, searching can still be carried out effectively on the **date written** field as a text data type.

6.2 Successes and innovations

The main success of this project was in the production of a usable finding aid from scratch, with no precedent within the Rugby School archives or for PC-based relational databases in the archival literature. What archival computerization projects have appeared in the professional literature have mostly involved large repositories utilizing mainframe computers to mimic existing manual practices. The biggest innovation in this project was the development of a computerized solution on its own terms, not the mechanisation of a pre-defined manual process or set of principles.

The infrastructure is now in place for the expansion of the finding aid to the remainder of the Arnold Papers and, subsequently, to other collections within the school archives.

6.3 Recommendations

As the infrastructure covers an incomplete collection, i.e. only the first five of the nine volumes of the Arnold Papers, there is a good opportunity for familiarization with the application by using the framework provided by this finding aid to complete the description of the Arnold Papers.

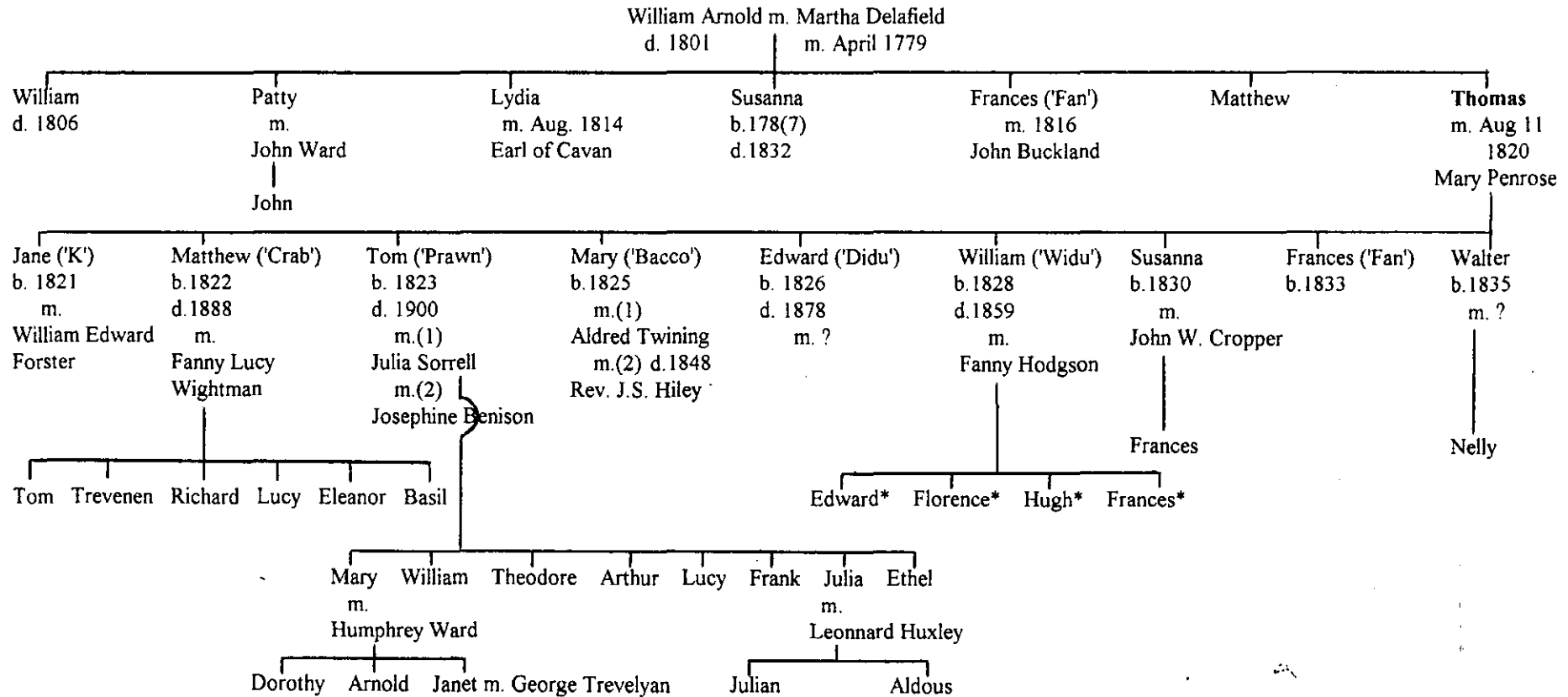
The learning opportunities afforded by completion of the Arnold Papers description would be very useful as a stepping stone to the production of another finding aid for a different collection. Using the principles of data gathering and analysis, database design and retrieval methods outlined above the next stage should be the expansion of these ideas to another collection, perhaps incorporating the indexes of different collections as separate tables but in a single consolidated database. Fundamental to the expansion of the infrastructure is a knowledge of the capabilities and limitations of the software used.

6.4 Conclusion

I came to this project as a computer professional coming to a problem of archival description, rather than in the more normal manner as an experienced archivist coming to computerization steeped in the principles and practices of my craft. Whilst readily admitting my own lack of experience and my great debt to the authoritative writings of distinguished archivists from Jenkinson to Cook, I feel that the novelty of my perspective on the problem has allowed me to blend my own knowledge of the principles of computerizing manual processes with the knowledge acquired during the project of the fundamental principles of archival description to produce a solution which solves the perceived problem in a way which complements the strengths of the computer, rather than using the computer to simulate a set of procedures, in terms of cataloguing, listing and indexing, largely designed to accommodate paper-based descriptions.

As far as the objectives listed in section 1.3 are concerned, I have proven by my examples in chapter five that the Arnold Papers database is a usable and useful finding aid. The third objective was to realize the advantages of computerization for data manipulation. Again, the examples in chapter five demonstrate that this has been achieved. The index tables of PERSONS, PLACES and SUBJECTS contain some 2364 records between them, from 253 documents. Whilst being reluctant to profess comprehensiveness for these indexes, the detail is clearly great. Finally, the sum of these objectives was to provide the librarian, pupils, and authorized researchers at Rugby School with a flexible system which will stand as finding aid, record control tool and model for expansion for some time to come. This, I submit, has been done.

The Arnold Family Tree



* Later adopted by Jane and William Forster, taking the surname Arnold-Forster

Bibliography

- Anderson, J. *Rugby School archives: The Arnold Papers*. Unpublished M.A. dissertation, Loughborough University of Technology, 1988.
- Armitage, Janet E. and Michael F. Lynch. *Articulation in the generation of subject indexes by computer*. University of Sheffield, 1967.
- Bamford, T.W. *Thomas Arnold*. The Cresset Press, London, 1967.
- Berner, Richard C. 'Manuscript collections and archives - a unitary approach'. *Library Resources and Technical Services*. Vol. 9, 1965, pp. 213 - 220.
- Bernier, Charles L. 'Subject-Index production'. *Library Trends*. Vol. 16, No. 3, January 1968, pp. 388 - 397.
- Chalmers, Duncan. 'Computer indexing in the Public Record Office'. *Journal of the Society of Archivists*. Vol. 6, No. 7, April 1981, pp. 399 - 413.
- Cook, Michael. *Archives administration*. Folkestone, 1977.
- Cook, Michael. *Information management and archival data*. Library Association, London, 1993.
- Cook, Michael and Margaret Procter. *A manual of archival description*. 2nd. ed., British Library, 1989.
- Cox, Richard J. *Managing institutional archives*. Greenwood Library Management Collection, Westport CT., 1992.

- Darwin, Kenneth. 'The use of the computer in indexing records'. *Journal of the Society of Archivists*. Vol. 4, No. 3, April 1971, pp. 217 - 229.
- Gracy II, David B. *Archives and manuscripts: Arrangement and description*. Society of American Archivists, Chicago, 1977.
- Green, Adam. 'Reference codes and computer-based systems'. *Journal of the Society of Archivists*. Vol. 11, No. 4, October 1990, pp. 144 - 146.
- Jasenas, Michael. 'Cataloging small manuscript collections'. *Library Resources and Technical Services*. Vol. 7, 1963, pp. 264 - 273.
- Jenkinson, Hilary. *A manual of archive administration*. Percy Lund, London, 1937.
- Kitching, C. *The impact of computerization on archival finding aids: a RAMP study*. UNESCO, Paris, 1991.
- Muller, S. J.A. Feith and R. Fruin. *Manual for the arrangement and description of archives*. Netherlands Association of Archivists, New York, 1968.
- Pepler, Jonathan. 'The impact of computers on classification theory'. *Journal of the Society of Archivists*. Vol. 11, No. 1, January 1990, pp. 27 - 31.
- Phillips, C.M. and C.M. Woolgar. *Computerising archives: some guidelines*. Society of Archivists, Winchester, 1985.
- Schellenburg, T.R. *The management of archives*. Columbia University Press, London, 1965.

- Scrase, Miranda H. *Creating an inventory: A finding aid to the records of the Derbyshire branch of the National Farmers' Union*. Unpublished M.A. dissertation, Loughborough University of Technology, 1991.
- Simpson, Alan. *Understanding Microsoft Access*. Sybex, Alameda CA., 1993.
- Society of Archivists. *Preparation of finding aids*.
- Storey, Richard A. 'Indexing archives'. *Indexer*. Vol. 5, No. 4, Autumn 1967, pp. 159 - 168.
- Taylor, Hugh. *The arrangement and description of archives*. International Council on Archives, ICA handbooks series volume 2, London, 1980.
- Trevor, Meriol. *The Arnolds*. The Bodley Head, London, 1973.
- Underwood, M.G. 'A computer index for the archives of St. John's College, Cambridge: a progress report'. *Journal of the Society of Archivists*. Vol. 6, No. 4, October 1979, pp. 214 - 218.
- United Nations Educational, Scientific and Cultural Organization. *Indexing principles*. Paris, 1975.
- Williams, C.J. 'Computerisation in the Clwyd Record Office'. *Journal of the Society of Archivists*. Vol. 10, No. 2, April 1989, pp. 66 - 74.
- Woolgar, C.M. 'The Wellington papers database: an interim report'. *Journal of the Society of Archivists*. Vol. 9, No. 1, January 1988.

