ONLINE ACCESS TO NATIONAL ART LIBRARY DOCUMENTS

An Interactive Qualifying Project Report

Submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfilment of the requirements for the

Degree of Bachelor of Science

by

_____

Adam Brancato

_____

Michael Modisett

_____

Alex Tang

April 27, 2001

London Project Centre

Project # RLS-LD15

Approved:

_____

Professor Ruth Smith, Co-Advisor

_____

Professor Jim Demetry, Co-Advisor

Authorship

AB = Adam Brancato, AT = Alex Tang, MM = Michael Modisett

	Primary Author – Primary Editor
Authorship	AT – MM
Abstract	MM – AB
Executive Summary	MM – AB
Table of Contents	AT – AB
1 Introduction	MM – AT
2 Background Research	
2.1 History of the Victoria & Albert Museum	AT – MM
2.2 History of the National Art Library	MM – AT
2.3 Archiving	AB – MM
2.4 Databases	MM – AB
2.5 Programming Languages	AB – AT
2.6 Human Computer Interaction	MM – AB
2.7 Library Technology	MM – AT
3 Methodology	
3.1 Overview of Objectives	AT – AB
3.2 Gaining an Understanding of the Researcher	MM – AT
3.3 Exploring the Functionalities of the XML Tags	AB – MM
3.4 Designing the System	AB – MM
4 Results & Analysis	
4.1 Analysis of NAL Patrons	MM – AB
4.2 Analysis of Archive Researchers	MM – AB
4.3 Analysis of TEI Tag Set	AB – AT
4.4 Analysis of EAD Tag Set	AB – AT
5 Conclusions & Recommendations: System Design	
5.1 Search Field Parameters	AB – MM
5.2 Database Table Design for Text Searching	AT – AB
5.3 Parser Design	AT – MM
5.4 Interface Design	MM – AB
5.5 Accessibility Issues	AT – MM
5.6 Continued Expansion of the Online Library	AB – MM
6 Bibliography	MM – AB

Abstract

The National Art Library (London) is currently transcribing manuscripts into an electronic format to better preserve and present the information contained within, and make the documents available for various research uses. Through surveys and interviews with researchers and Library staff we designed an online text search and display system for these transcriptions. Our design includes recommendations for a database, parser, interface, and universal accessibility options, as well as a prototype to demonstrate the system's functionalities, all with an emphasis on human-computer interaction.

Executive Summary

The National Art Library (NAL), housed in the Victoria & Albert Museum (V&A) in London, contains numerous priceless manuscripts documenting the history of art and the Museum. Unfortunately, researchers who wish to use these documents face a number of obstacles. Due to their age the documents are deteriorating and becoming difficult to decipher. In addition, researchers must travel to the NAL in order to view them and upon arrival access is further hindered by the tight security and seating limitations in the Library's reading room. To help alleviate these problems, the Library is currently transcribing some of the most valuable and useful documents into a variety of electronic formats, one of which was designed by a previous Worcester Polytechnic Institute students' Interactive Qualifying Project team.

The previous IQP team designed a tagging system that allows transcribers to mark the text of the documents with XML tags. These tags annotate the text, separating important information out so that it can be easily recognised by a computer. The Museum is using two tag standards for a variety of documents: the TEI tag set and the EAD tag set. The TEI, or Text Encoding Initiative, tag set is used for documents where the full text is available. The previous IQP group modified this tag set to encode art history documents, specifically the Robinson Reports. The Museum staff at the V&A Archive at Blythe House is using the EAD, or Encoded Archival Description, tag standard to encode archival abstracts that are housed there. The EAD tag set is another standard that was designed specifically for abstracts, and contains specialised tags for abstracted information. The goal of our project was to design the specifications for an online library resource that would make the text of these TEI and EAD encoded documents available over the internet with searching and

embedded linking capabilities, thus providing quicker, easier and more powerful access to these priceless resources. Because these documents are essential to NAL patrons, human-computer interaction research was crucial as it provided us with important design concepts to help make the information within the electronic texts easy to access.

The first step in reaching this goal was to gain a sound understanding of the research community that would be using this system. Through surveys and interviews we determined the archive documents that researchers use the most, the reasons they use them, the other sources they use for similar information, and the type of functionality that would be most useful in the online system. This was a crucial stage in our methodology because in order to design the most appropriate system for the NAL and its patrons, we needed to understand who would be using the system and why.

After gathering information about the researchers, we studied the technical concerns about the programming and design elements that could be used to create an online text search and display system of this type. We studied a variety of technical subjects including database layout, parser design, TEI and EAD tag standards, universal search protocols, and human computer interaction issues. By combining our technical knowledge as engineers with our newfound understanding of the research community at the NAL, we were able to design the specifications for a comprehensive and flexible online library resource.

The system we designed is straightforward to use and easily extendible for future transcriptions and tagging methods. We made sure to design an interface similar to the online library catalogue that is already available in order to minimize the amount of time needed for researches to learn how to use the online text program.

The main parts of the proposed system include both a simple search page to accommodate new users as well as an advanced search page for those already familiar with the documents. The search results page is similar to that found on most major internet search engines and includes information about the title and date of creation as well as a few lines of the text of a document. The text display page is arguably the most important part of the system and thus is very intuitive and malleable based on each user's preferences. This page also contains links to online resources outside of the NAL, thus increasing the gateways to information available to the users. Our design decisions and recommended program functionality are summarized in a prototype of the system (Appendix E).

All of these design specifications are meant to meet the desires of the researchers and potential users of this system while still meeting the NAL's criteria of low cost and ease of implementation. Once this system has been implemented, the efficiency with which scholars will be able to research the online documents will be significantly improved. This system will provide new ways in which scholars will be able to use these extremely important and interesting sources of information and will make them accessible to a wider range of scholars and organisations. As more collections are made available electronically, the range and efficiency of scholarly research will be greatly increased. Furthermore, the online library resource that we designed is extremely flexible and may be transferable to similar art history archives around the world.

Table of Contents

1 Introduction

The National Art Library (NAL) is housed in the Victoria & Albert Museum in London, England. During the one hundred and forty years that the museum has been open, the NAL has collected over two million items, including several original manuscripts and personal papers. These documents represent a great deal of history, and most of them are one-of-a-kind and irreplaceable. Some documents are very old and have become both illegible and extremely fragile; therefore, restoration and preservation techniques are continually being used to extend the lifespan and usability of these documents. Currently, a person who wishes to use the NAL's resources must travel to London to study these papers. A previous Worcester Polytechnic Institute students' Interactive Qualifying Project group designed a tagging system so that selected documents could be transcribed and stored electronically (Holt, Kiffer & Peterson, 2000). Since that time, NAL staff and outside researchers have begun translating various original documents, specifically the Robinson Reports, Art Referees' Reports, Abstracts of V&A Correspondence, and Board Minutes of the Science & Art Department, into electronic format. The goal of our project was to design the specifications for an online library resource: a method to display the text of these documents, as well as documents that will be transcribed in the future, over the internet. Once our design has been implemented by another party, it will be possible for anyone with web access to view these historic documents online through the NAL website.

There are several important benefits to providing online access to electronic versions of these documents. First of all, online access means quicker and easier access for many researchers because the internet can reduce or eliminate the need to travel to the NAL to view the documents first hand. In addition, multiple people can

read documents online at once, whereas only one person can use original manuscripts at a time. Due to space limitations, seating in the NAL is at a premium and security passes to view some of the most rare documents can be hard to come by; online texts eliminate the need for both of these.

Online texts provide not only faster and easier access to information, but also more powerful access through added functionality. Searching tools allow researchers to find what they are looking for quickly, and may point them in new, previously un-thought-of directions as well. Links embedded within the transcribed text to related documents or pictures of art objects housed online provide yet another source of information. Embedded links can also point researchers to resources outside of the NAL such as university libraries, government archives, personal home pages or corporate websites; the list of potential links is endless, which is one of the most powerful features of an online library system and the internet in general.

Another key benefit of storing documents in an electronic format is the inherent ability to reformat computer text to an individual's desire; double spacing, highlighting or striking-out text are just a few of the flexible presentation techniques that are available to computer users that allow easier viewing and printing of information. Clearly, an online library system provides researchers with quicker, easier, more individualised, and more flexible access to information and can greatly enhance the research process.

A major part of designing this online resource was determining page layouts that researchers would be comfortable with, as well as the types of information that they wish to search for in the documents. It is important that the search methods are comprehensive, because the computer should be able to find at least as much information as a researcher would by reading these documents, in essence creating a

valuable timesaving device. Our result was a user-friendly online search system that provides enough keyword information for researchers to find desired documents and view the text in a web browser. We submitted to the NAL design specifications for a database, parser, interface layout, as well as a prototype of the system. We also discussed the options for XML display methods and universal accessibility through standardised protocols.

Because libraries and museums throughout the world are examining ways to provide online access to their resources, our findings could be of great interest to librarians, researchers, and museum curators. Once the NAL creates a working model of this new system of information access, any person with internet access will be able to view the information that is currently stored in the Library without ever entering the NAL. Other museums and libraries are also working towards a similar goal of creating a comprehensive online library for their material, but no one has been entirely successful. The work we did has global implications because once our design has been implemented, libraries around the world will be able to use the NAL as an example of how to create an online library of their own.

In order to achieve these goals, we created a number of preliminary objectives for our project. First, we established an understanding of the NAL, its history, and most importantly its users. In order to determine what resources they use and how online information access would be of benefit to them, we surveyed and interviewed researchers who use the NAL and have used the archives that would initially be placed online. We also determined the features they would like to see in the online document display system, such as the possible inclusion of embedded links within the text. In addition to these interviews, we also studied a survey conducted last year by the NAL to learn about general trends and expectations among Library patrons. After

gathering information about the researchers, we studied the technical concerns about the programming and design elements that can be used to create an online text search and display system of this type.

We used the information we collected to design the online system that meets the needs of the NAL and its patrons. It included easy-to-use yet comprehensive search functionalities in order to assist researchers with basic information gathering. Our online library design was the result of balancing the researchers' needs with the NAL's computer capabilities and we feel that once implemented, the system will be extremely useful to those doing research with NAL archival documents.

In order to familiarise themselves with the relationship between society and technology, Worcester Polytechnic Institute students participate in an Interactive Qualifying Project (IQP). Our project is about designing a system that will serve a specific community, the researchers that use the NAL. However, it is possible that once these documents are more accessible, they will gain more widespread use throughout the world. Therefore, we designed the specification for a technology that will be used by researchers, but perhaps also by the general public.

Over the past few years, as much of the world has become accustomed to the amount of information available online, museums have had to reassess their traditional methods and goals. People now have the ability to call up information easily with a computer and find what they are looking for by "surfing the web." Because the internet has changed the way that society communicates and even thinks, the usually cautious and reserved museum communities have had to adapt to meet the new public perceptions of how information should be made available. As current issues and ideas are placed online through media outlets, company websites, and personal homepages, it is extremely important for historical information to be

represented in the digital world as well. For many, if history does not exist online, then ideas of the past can become lost in inaccessible archives. Museums and libraries represent an important link to the past, and that link must move forward if it is to continue to play an important role in society.

2 Background Research

2.1 History of the Victoria & Albert Museum

The NAL is housed in the Victoria and Albert Museum (V&A), which was founded in 1852, and originally titled The Museum of Manufactures. Its founder, Henry Cole, stated the mission of the Museum in his first annual report to the Board of Trade:

> "[A] Museum presents probably the only effectual means of educating the adult who cannot be expected to go to school like the youth, and the necessity for teaching the grown man is quite as great as that of training the child. By proper arrangements a Museum may be made in the highest degree instructional. If it be connected with lectures, and means are taken to point out its uses and applications, it becomes elevated from being a mere unintelligible lounge for idlers into an impressive schoolroom for everyone" (Baker, 1999, p. 9).

The museum was soon renamed The South Kensington Museum and remained this way until 1899 when Queen Victoria gave it the current title in her last public appearance as a tribute to "her beloved consort Prince Albert" who had died 38 years earlier (Baker, 1999, p. 11). From its inception, the South Kensington Museum, "though benefiting from some outstandingly generous private bequests, contained collections that either had been purchased with government funds from… international exhibitions, or assembled for the use of the Government Schools of Design" (V&A, 2001).

Over the past century and a half, the museum has managed to assemble an impressive and extensive collection of applied arts, which have been arranged into eight major categories. The first is the ceramics and glass collection, "which can claim to be the most important in the world for the area it covers" (V&A, 2001). Second is the Far Eastern Department, which contains items from China, Japan and Korea dating from as early as 4000 BC. In addition, the V&A houses the foremost

collection of furniture and woodwork in the world, containing over 14,000 pieces from Britain, Europe and America dating as far back as the Middle Ages. The Indian and South-East Asian Department also contains an impressive collection of applied art from India, Pakistan, Bangladesh, Sri Lanka, Afghanistan, Tibet, Nepal, Bhutan, Sikkim, Burma, Thailand, Indonesia and Malaysia. The Metalwork, Silver and Jewellery Department houses over 45,000 examples of decorative metalwork, some of which date back as far as the Bronze Age. Paintings were some of the first works to enter the museum, and the V&A's current "collection is unrivalled, including such masterpieces as the Raphael Cartoons, on loan from Her Majesty The Queen, and every day graphics such as the London Underground map. Nowhere else in Britain can the representation of the three-dimensional world in two dimensions be studied in such depth and variety" (V&A, 2001). The Sculpture department houses the national collection of post-classical European sculpture, and contains pieces largely from before the First World War. Finally, the textiles and dress department covers a period of more than 2,000 years and includes a broad spectrum of examples from across the globe, with particular emphasis on Europe (V&A, 2001). Combined, these departments hold more than 4.3 million objects – truly the V&A is one of the world's most impressive museums.

2.2 History of the National Art Library

The National Art Library (NAL), located on the second floor of the V&A Museum, is "the largest and finest art library in the world" (Banks, 1973, p. 450). The NAL is both a major reference library and the Victoria and Albert Museum's curatorial department for the art, craft and design of the book (NAL, 2001). Subjects covered by the Library include those central to the work of the Museum as discussed above, but the "fundamental strength [of the library] lies in the range and depth of its

holdings of documentary material concerning the fine and decorative arts, from many countries and periods" (NAL, 2001). Materials for the Library are acquired in most Western European, as well as some Asian, languages. There are no restrictions by date, and a wide range of formats is collected, from manuscript material to videodiscs and CD-ROMs. Numerous private libraries have been donated to the museum over the years, greatly adding to an already impressive collection. Of particular note are the collections of calligraphy and 20th century book art; the Dyce Collection, strong in the theatre, literature and the classics; and the Forster Collections, strong in literature and history, with many manuscripts and a large collection of 19th century pamphlets (NAL, 2001).

Our project focused on several types of archival documents that the NAL has collected throughout the years, and it was important to understand the purposes for which these archives had been created so that we could design an appropriate search system. As the V&A is a museum dedicated to the collection of artefacts pertaining to popular culture as well as more traditional collections of sculpture and paintings, the NAL has collected books, papers, and personal correspondences unlike any others in the world. Archives such as the Robinson Reports, the correspondence abstracts, and minutes of board meetings since the opening of the museum were some of the documents that we worked with. Understanding the methods of archiving and cataloguing was extremely important to accomplishing our goal.

2.3 Archiving

Cataloguing an archive is one of the most important tasks that an archival custodian has, because cataloguing provides the researchers with a link to the information within an archive. The archive system must be orderly and concise, in order to ensure that none of the information becomes lost. Standards for archiving

have been developed both worldwide and specifically within the United Kingdom (Christopher Marsden, personal communication, 3/15/2001). These standards are used to organise a collection in a set way, so that patrons will have a system that they are familiar with whenever they enter a new library. This system, while standardised, allows a library to specify the terms and classifications that it uses in order to serve the users and satisfy the particular terminologies that might arise from the documents that it contains.

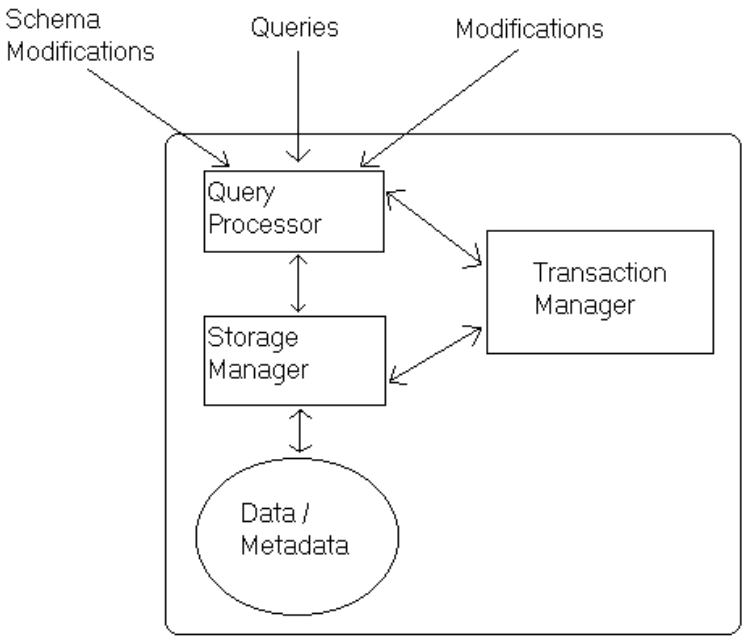
The Robinson Reports have been reorganised at least three times since they were originally entered into the archive (Fontanella, personal communication, 2/09/2001). They are currently organised by the NAL's official designation system, which all documents within the NAL possess. The previous IQP group that worked at the NAL set up the tagging and cataloguing system that was used to transcribe the Robinson Reports (Holt, Kiffer & Peterson, 2000), which is now one of the templates for all future electronic documents that the NAL will put online. The electronic format will allow for searchability, so the team maintained this ordering for labelling the files, because the documents will be searchable by any number of categories. The original order is listed in order to accommodate people that wish to see the actual paper and not just the electronic transcription. They chose to focus on the Text Encoding Initiative (TEI) electronic organisation and to name the electronic documents following the last organisational system that the NAL had set up for the Robinson Reports.

The Text Encoding Initiative (TEI) is an organisation working to standardise a means of representing text-based information (documents, papers, books, etc.) in electronic format. The main functions are to determine what text should be encoded and to create a system to prevent information loss on different operating systems, and

communicating between different operating systems (Electronic Text Center at the University of Virginia, 2001, Preface, p. 1). The TEI system's standards were applied in creating the tag system for the Robinson Reports. Using the TEI standard means that the search parameters contains similar types of terminology to other systems around the world, while still allowing terms specific to the types of documents that the NAL wants to put online. These TEI tags are the tools that allowed us to set up the search system for these documents, by using a database to store the information contained within the tags. This database then organised the information internally to provide quick text-search options.

2.4 Databases

Simply put, a database is a collection of information that is managed by a database management system or DBMS. A DBMS is expected to embody four major concepts. First, it should allow users to create new databases using computer languages called "data-definition languages." Second, a DBMS must allow users to query the database, meaning the users must be allowed to ask the database questions as well as modify the contents of the database using another language called a "data-manipulation language." A DBMS is also required to be able to store information, usually over a long period of time. Finally, a DBMS is expected to control access to the database to increase security and reduce simultaneous manipulation errors (Ullman & Widom, 1997, pp. 1-2). To fulfil these requirements, a DBMS is composed of several components as illustrated in the following diagram adapted from Ullman & Widom (1997, p. 7):



In this diagram, the four major parts of the DBMS are located in the large square: the data/metadata, query processor, storage manager, and transaction manager. The area in the circle indicates where data are stored (usually on hard disks or some form of optical media.) In addition, most databases contain information about *how* the data are stored. This is referred to as "metadata" and defines the organisation of the information. The storage manager is responsible for obtaining information from the data storage and modifying the information there when requested to by the levels of the system above it (Ullman & Widom, 1998, p. 8). It is the transaction manager's job to ensure integrity of the system. In the context of databases, "integrity" refers to the security of the system as well as the guarantee that data will not be lost in case of a system failure. The query manager handles input to the database, usually submitted in some fashion by the user.

2.4.1 Inputs to a DBMS

As indicated in the diagram, there are three main types of inputs to the DBMS: queries, modifications, and schema modifications. Simply put, queries are questions

about the data. For example, a librarian could request the number of books currently overdue and the DBMS would access the appropriate database and return the requested information. It is important to note that a query can never change the contents of a database. To handle this, a DBMS must also be able to accept modifications. For example, the bank teller could modify the database by increasing the amount of money in a certain account. In addition to modifying the information stored in the database, the system must also allow the user to modify the way the information is stored. These types of modifications are called schema modifications and are used to reorganise or add new fields to the database. For instance, a bank may want to add a new area of information to be stored in the database such as the client's email address. Because it would not modify existing data, this task could only be accomplished through a schema manipulation.

2.4.2 Data-definition Languages

There are currently two standards for data-definition languages: *the object definition language (ODL) model* and the *entity-relationship (E/R) model*. Each has its advantages and disadvantages.

2.4.2.1 The ODL Model

The fundamental building block of an ODL definition is the object, which is composed of three properties: attributes, relationships and methods. Attributes are "properties whose types are built from primitives such as integers or strings" (Ullman & Widom, 1998, p. 28). Similar to spoken languages, attributes are used to describe objects. The second element of ODL objects is relationships. Again, much like in spoken languages, a relationship is essentially a reference to some other object. Finally, methods are "functions that may be applied to objects" (Ullman & Widom, 1998, p. 8). For example, the following is an example ODL definition for an object

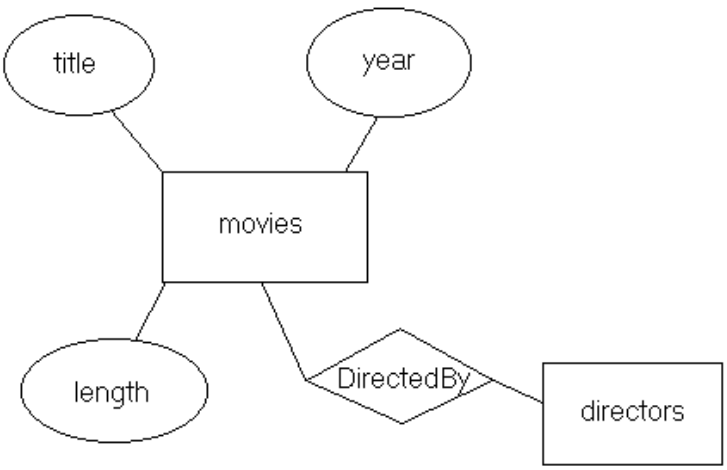
'movie' with attributes 'title', 'year' and 'length', as well as a relationship with 'directors':

Interface *movie*{
Attribute string *title*;
Attribute integer *year*;
Attribute integer *length*;
Relationship DirectedBy<Directors> *director*;}

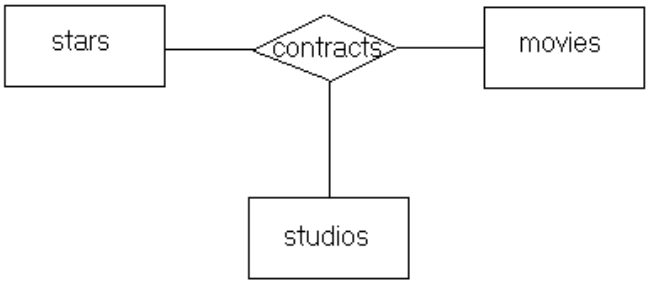
ODL is an extremely useful tool because it allows "object-oriented designs of databases to be written and then translated directly into declarations of an object-oriented DBMS" (Ullman & Widom, 1998, p. 26). This means that turning an ODL model into an actual implementation is fairly simple. However, the major drawback to the ODL model is that it is more difficult to create and understand than the E/R model.

2.4.2.2 The E/R Model

The E/R model is "the more traditional approach" to database design and is graphical in nature with boxes and arrows representing data elements and their connections respectively (Ullman & Widom, 1998, p. 25). E/R diagrams are similar to ODL definitions in the sense that they both have the same three main properties: entities (which are analogous to objects in ODL), attributes and relationships. However, the actual declaration of a database in the E/R model is quite different than in ODL. The following is the same 'movie' example as above, except created with the E/R model:



The main advantage of using E/R diagrams to model a database is that they are much easier to create and understand at a quick glance than ODL definitions. In addition, E/R diagrams provide added functionality that is not possible in the ODL model. For instance, there is no equivalent ODL definition for the following:



However, there is one main drawback to using E/R diagrams: when it comes time to translate the model into an actual implementation of the database, the programming is much more difficult. The E/R model is easier to understand due to its graphical nature, but this is the same feature that makes it so difficult to programme.

2.4.3 Querying and Modifying Databases

While the ODL and E/R models are useful for the high-level organisation and conceptualisation of databases, an actual programming language is needed to make use of the real-world database. "The most commonly used database systems query and modify the database through a language called SQL (Structured Query Language), also known as 'sequel'" (Ullman & Widom, 1997, p. 243).

2.4.3.1 Querying Databases

All SQL queries are structured as follows:

SELECT	{list of attributes matching the conditions given in 'from' and 'where'}
FROM	{list of places to look for the above attributes in the database}
WHERE	{list of constraints on the attributes in question}

For example, to tell the DBMS to return the titles of all movies made in 1980 that were shorter than 120 minutes, one would type the following SQL command:

```
SELECT    title
FROM      movies
WHERE     year = 1980 AND length < 120
```

2.4.3.2 Modifying Databases

Unlike queries, database modification statements do not return a result but rather "change the state of the database" (Ullman & Widom, 1997, p. 279). There are three types of modification statements that allow one to insert, delete or update values in a database. Insertion statements are used to add new information to the database. Insertion statements follow the form:

> INSERT INTO {where to place the new information} VALUES {raw data to insert}

Deletion statements are used to remove existing information from the database. Deletion statements follow the form:

> DELETE FROM {where to delete the information from} WHERE {conditions that define which data to delete}

Update statements are used to change existing information in the database. Update statements follow the form:

> UPDATE {where in the database to update the information} SET {which attributes to update} WHERE {conditions that define which data to update}

2.4.4 Databases in Libraries

While the design and implementation of databases are relatively simple, the effects of a well structured, thorough and easily accessible database are numerous and profound. Any large collection of information should almost always be stored in the form of a database so that it can be catalogued, organised, modified and searched easily. Libraries have been using databases for decades with the card catalogue

system. Recently, with the proliferation of computer resources, libraries have started putting their card catalogues in digital databases to increase searching capabilities. Due to the rapidly expanding expectations of users, libraries will soon be making these databases available to the entire world via the internet.

2.4.5 Databases in the NAL

In 1999, the NAL made their card catalogue available online. This system uses a database to store entries for each item in their collection – each item has its own entry. Within this entry are fields for title, author, publisher and other similar information. When the user searches for library materials with this system, the search engine reads through the database and returns to the user a list of materials that match the given criteria. The user then selects one of these criterion and the system displays the database entries for that item, i.e. the title, author, etc. This system works very well for this application, but we plan to use a database in a slightly different manner for the online text system.

All of the transcribed documents will be placed on a computer in a standard file directory system. Then a parser will read through these documents and extract the contextual keywords that were marked by the transcriber. A database will be used to store these keywords, along with links to their matching electronic text documents. In other words, the documents themselves will not be in a database, only the keywords will. This database structure provides a number of important benefits to both the Library and the user. It helps the Library because it will be easier to maintain, add new documents, alter existing documents and keep track of where the documents are electronically located. It will also allow the NAL to easily extend the functionality of the system by coding new tags into the DTD file. Finally, if someone ever tries to hack into the system, accessing the actual transcribed documents will be extremely

difficult. This way of using the database will also be of benefit to the user because it will run faster and allow greater concurrent access to the system.

2.5 Programming Languages

An understanding of computer languages that are used in the creation of webpages and online systems formed the basis for the specifications of the online system because a full understanding of the capabilities of a computer allowed us to utilise those capabilities to their fullest potential. The NAL has some documents stored in SGML format, while the Robinson Reports and several others are in XML format. In addition, webpages themselves function using HTML and sometimes also contain a Java application. Each type of markup language has strengths and weaknesses, suiting them to different tasks.

2.5.1 SGML

The first markup language was SGML, Standard Generalized Markup Language, the original standardised language to format the design of a document or webpage. SGML was an outgrowth of IBM research in the 1960's about making a markup language, which IBM called GML (General Markup Language). It was standardised over the next years until the first version of SGML was published in 1986 (Johnson, 1999, p. 1). All SGML formatting was done by creating tags within the document, pieces of code that define every aspect: text size, justification, background and text colouring, all aspects of the design. Programmers could define their own tags for their documents and had complete control over the layout and structure of the page. "However, full SGML contains many optional features that are not needed for Web applications and has proven to have a cost/benefit ratio unattractive to current vendors of Web browsers" (Bosak, 1997, p. 2). SGML is so complex that coding a document can take a large amount of time, and it is difficult for

a person to use it to code a simple document, without all of the extra functions that are possible. So an extremely simplified and refined version of SGML was created, called HTML.

2.5.2 HTML

HTML is a markup language that is very structured and limited to presenting simple data, designed to give instructions on how to format a page (Johnson, 1999, p. 1). It is not a language that can be used to compute information, only display the information. Most webpages are written in HTML code, because it is so simple to use. However, while it is an excellent language to use for basic design, it has many shortcomings that would trouble an advanced user. Johnson (1999, p. 1) lists a few of these deficiencies. Primarily, HTML is a set language, not extensible. Users cannot create their own tags in order to organise data in a certain way; only tags that were initially included in the browser will function. HTML also provides no way to display the data in a different format from the original layout based on user preferences. The initial layout created by a page designer is the only format that a viewer can use. Finally, HTML cannot define the semantic meanings of a word. A tag cannot be searched for one definition of a word; it will find and display all instances of the word, unable to check for the validity of the context. Bosak (1997, p.1) also notes another flaw, the simplicity of HTML. It cannot be used to create complex structures to represent databases or "object-oriented hierarchies." Because HTML has all of these weaknesses, a better markup language was designed.

2.5.3 XML

XML is another markup language, but unlike HTML, "…XML is a *metalanguage* -- a language used to define new markup languages. With XML, you can create a language crafted specifically for your application or domain" (Johnson,

1999, p. 1). This markup language gives a programmer freedom from the constraints of the more-structured HTML. Jon Bosak (1997, p. 2), one of the original creators of XML, describes the main advantages of the new language. XML allows page designers to create new tags that are needed for a document. The document structures can also be more intricate or complex if such a nested structure is called for. And finally, XML can contain grammatical descriptions of tagged words if such definition is required. XML is a much more functional language, returning flexibility to page design without regressing to the complications of SGML.

One of the features of XML that we were concerned with was the ability to nest information within the tags themselves. For example, you can include within tags extra contextual information about the words that have been tagged. The previous group created several tag subcomponents, which the transcribers can use to add additional information to the document. This information will not normally be visible since it was not part of the original text, but the computer will know that the information is there, and that people can access it if they so desire.

It is important to note that in these documents there are two types of tags, formatting tags and contextual tags. While both are important, we are concerned mainly with the contextual tags. The formatting tags specify only the physical layout of the documents and how they appear on screen. The contextual tags contain information about the document such as names and dates, marking these for the computer to notice. All documents need to be tagged completely for format and for content in order to make the search and the display of the electronic transcriptions work as desired.

2.5.4 Parsing the Data

To create this type of search system, once all of the XML documents have been tagged, the tags must be sorted into the database. This can be done by hand or it can be performed using a parser. Generally speaking, a parser is a piece of software that scans through electronic files and extracts words or phrases that the user asks it to. For XML, parsing means splitting the document by its tags, which will reveal the structure and attributes of the tags and the document as a whole (Sall, 1998, p. 2). For our project, we used a parser to comb through the electronic text of the XML documents and extract the tags and the tagged text and then place this information into the appropriate tables of the database. Each time a new document is transcribed and added to the collection of electronic texts, the system will have to run the parser so that the tags from this new text are added to the database. Only after parsing will the text of a document be available to users of this online system. The parser therefore had to be compatible with both the database and the XML documents.

2.5.5 XSL Style Sheets

Once the documents have been tagged, it also necessary to set up a system to view them in their formatted form without tags cluttering up the screen. For a markup language such as XML, this means using a type of style sheet. A style sheet allows a programmer to specify the particulars of the display the tag elements. Once all of the information has been tagged, the style sheet will display the information in the formatting specified (Johnson, 1999, p. 4). XSL is the style sheet that the NAL has chosen to use for its documents. XSL is a multifaceted style sheet with numerous functions. In addition to formatting the document for display purposes, XSL sheets can translate XML into HTML, or into a different XML dialect. XSL is an extremely powerful tool when working with XML.

One problem with using XML is that older web browsers are not configured to view XML or XSL directly. In order for these browsers to view the information, either the XML must be translated to HTML for display, or a companion programme of some sort must be created to expand the capabilities of the web browsers. Both of these options are feasible for the NAL, it is simply a matter of choosing the route that they wish to follow.

2.5.6 Java

One method of online information display is the use of Java. Java, created by Sun Microsystems, is not a markup language, but a programming language. If used in a webpage, a Java programme is downloaded from the site and then runs on a computer. Because it is a programming language, it is more complicated to create the code that is needed. However, it provides a whole new array of functionalities to use on a page. For example, the current NAL webpage (NAL, 2001) has a Java-based catalogue that can be used to search for titles of documents based on keywords. Java usually functions by creating a separate region within a page, where all of the programme's functions can be used. This programme is called an applet, which means a miniature application. "Java… [is] a good language for things like database front ends and other lightweight applications" (Shiffman, 1998, p. 1). The main advantage is that "Java code compiled on one kind of computer will run on every other kind of computer with a Java interpreter" (Shiffman 1998, p. 1). Java can be used to create interactive programmes that are portable, meaning able to work on different platforms. This is crucial for an internet programme, since all recent web browsers contain Java interpreters, all computers can view a Java applet.

Java allows more functionality than a markup language, because it is a true programming language. It is easy to integrate a style sheet to display XML within the Java applet, and Java can interface with a database if it is set up correctly. However, there are a few drawbacks to Java. Because the programme needs to be downloaded to the user's computer, it can take a lot of time and bandwidth to download a large and complex programme. Also, because the code is being downloaded to the user, a person can capture the code and break it down in order to understand its functions and probe for weaknesses, causing a potential security risk.

2.5.7 CGI

CGI, or Common Gateway Interface, is another type of programming language often used in online systems by creating CGI scripts to execute commands. CGI scripts can be made with any type of language, and can be used to create a programme to perform similar functions to a Java programme. However, rather than downloading the interface programme to the user's computer, the programme is maintained on the server. The user enters his or her query information into the HTML-based interface, the information is sent back to the server and processed, and the results are returned to the user. This means shorter download times for the user, because the programme is not downloading onto his or her computer, but there are a few drawbacks. Because the server is running the programme, a large number of users working at the same time will create too many instances of the programme, and can use a lot of the server's resources, which will slow down or crash the system.

2.5.8 JavaScript

JavaScript, although it has a similar name to Java, is a completely separate language. It is a scripting language, not a true programming language meaning that it does not have to be compiled on every computer it runs on. JavaScript is used to

perform simple tasks by a web browser, such as display instructions. Similar to Java, the JavaScript commands are downloaded to the user's computer, where they are immediately executed. The difference is that Java needs to be compiled in its entirety on the user's computer before the programme will run, whereas JavaScript can be executed one command at a time. JavaScript can be used to display more complex information than HTML, but will not provide as much functionality as Java. JavaScript is sometimes preferred over Java because, for simple tasks, it requires less code and therefore less download time. However, JavaScript cannot be used to make older browsers display XML by itself. It does, however, allow more functionality and flexibility with XSL, as long as the browser itself has the ability to display the XSL style sheets.

2.5.9 Perl

Perl is a programming language used for text handling and manipulation. It was very limited when it was initially created, but has evolved into a complex programming language capable of almost any task. Perl can be used to write a programme on the system server to perform various functions. For example, a Perl script can be written to translate an XML file and convert it to HTML, which can be displayed directly by a web browser. When using a Perl programme of this type, there is no need to store copies of the XML files in HTML format or to create Java or JavaScript programmes to be downloaded to users' computers to allow them to view the XML correctly. Instead, the Perl script will translate the XML into HTML instantly for display on the users' screen. This HTML copy will not be stored on the server anywhere, but rather only used that single time. This saves space and prevents creating redundant copies of the information, while still allowing access to the documents.

All of the computer components of an online system tie together to provide the user with an easy-to-use front end. This front end is called an interface, and recently the study of interfaces has become more important. This "human-computer interaction," or HCI is the study of what features and layouts make an interface enjoyable to use.

2.6 Human Computer Interaction

Human Computer Interaction (HCI) is "what happens when a human user and a computer system get together to perform tasks" (Hix & Hartson, 1993, p. 5). The field of HCI includes such diverse issues as user interface software and hardware, user and system modelling and cognitive and behavioural science. Despite the fact that these issues have been around for decades, HCI research and design as a profession is a relatively new concept that has largely arisen over the past decade. Early computer systems (from the 1950s to the mid 1980s) were extremely user-unfriendly. This is because relatively few people used computers during that time, and most users that did exist were members of large corporations or staff of academic or government organizations with a very high level of education. Designers of computer systems believed that an easy to use interface was unnecessary because the users of the system would be able to figure out how to use the system anyway. However, with the proliferation of computer resources and the increase of software functionality (and therefore complexity) in the 1990s, improvements in HCI were necessary. Computer professionals now realise that a good user interface is crucial for the success of any computer system because to users, "the interface *is* the system" (Hix & Hartson, 1993, p. 1). An easy to use interface is especially crucial for the NAL's online text system because its intended audience (primarily history and art researchers) is one that, as a

whole, has yet to master computing technology. The system needs to be as easy to use as possible in order to encourage the researchers to use it to its full potential.

2.6.1 Goal of HCI Studies

The goal of most work in the field of HCI is to provide the user with a high degree of usability, which is a combination of five user-oriented characteristics: ease of learning, high speed of user task performance, low user error rate, user retention over time and subjective user satisfaction (Hix & Hartson, 1993, p. 3). These are in contrast to computer-oriented characteristics such as computation time, strain on the processor, size of disk drives, complexity of computer code, etc. In sum, usability is related to the "effectiveness and efficiency of the user interface and to the user's reaction to that interface" (Hix & Hartson, 1993, p. 3).

2.6.2 Interaction Component and Interface Software

One of the key concepts in HCI is the *interaction component*, which is the way a user interface works – its behaviour in response to what a user sees and hears and does while interacting with the computer. In this sense, the process of designing a good interface is both a science and an art, yet it "draws on the engineering idea of making things good enough, but not perfect" – unfortunately, there is no such thing as the *perfect* computer interface (Hix & Hartson, 1993, p. 6).

Another key concept in HCI is the *interface software*, which is the actual implementation of an interface in some form of a computer language. Clearly, the interaction component must at least be acknowledged before work can begin on the interface software. An interface that is created without the use of the interaction component will almost certainly provide a very low degree of usability.

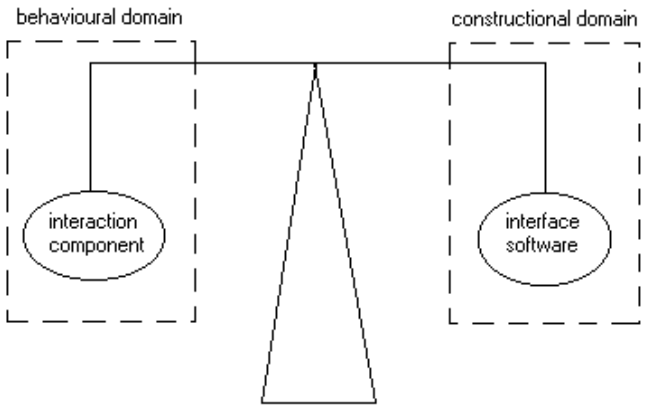
In a sense, this is one of the most important parts of our project. While we are not doing any of the actual implementation of the programme, we did create the

interaction component in the form of a Visual Basic prototype of the system. This prototype is intended for three audiences: the programmers of our design, current users of archives and future users of the online system, and the NAL. Because creation of the interaction component hinges on a good understanding of the users of the system, we decided to complete this work for the programmers. By doing so, we are ensuring that the interaction component will be acknowledged and the resulting system will be more user-friendly and provide a high degree of usability. In addition, by showing the prototype to archive users, they were able to suggest corrections and refine the system, resulting in a more complete, efficient and effective design. This is the basis of the principle of participatory design, discussed below. Finally, the prototype was very useful to the NAL because it represented a graphical summary of our design specifications of the system.

2.6.3 Behavioural and Constructional Domains

Because of the important "distinction between the development of the interaction component of an interface and the development of the interface software that implements that interaction," it is useful to think of two different domains where the respective work occurs: the behavioural domain and the constructional domain (Hix & Hartson, 1993, p. 6). The interaction component is designed in the behavioural domain where interaction is described abstractly in terms of the behaviour of the user and the interface as they interact with each other. This involves human factors such as cognitive limitations, graphic design, interaction styles and usability specifications. On the other hand, the interface software is developed in the constructional domain where software engineers develop the software to implement the behavioural design. This process involves computing factors such as algorithms, procedure libraries, data control, data flow, event handlers, error checking and object

oriented modelling. As HCI professionals know, "approaching user interface development in the behavioural domain, from a user and task view, should result in higher usability than approaching it from the constructional, or programmer's view, where software is the primary focus" (Hix & Hartson, 1993, p. 7). Unfortunately, an inherent conflict arises because what is best for the user is rarely best for the programmer. The final product is a child of these two parents, and the best interface will be the one with a balance between functionality and manageability.



For the NAL's system, we need to provide the highest degree of functionality to the user via searching and reading capabilities, yet make the database and parser easily manageable for the Library staff. Too much of either of these elements would tip the scale and render the system either non-functional or unmanageable.

2.6.4 HCI Guidelines

While the perfect computer interface does not exist, there are a number of guidelines that have been developed that can contribute to high usability in an interface. One of the best-known examples of a collection of guidelines is the compilation by Smith and Mosier at the MITRE Corporation (Hix & Hartson, 1993, p. 20). This collection of almost 1,000 guidelines evolved over more than a decade and

is extremely exhaustive in its coverage. What follows is a brief summary of some of the most important guidelines for creating a good user interface.

2.6.4.1 Practice user-centred design

User-centred design is closely related to the concept of behavioural design, producing the interaction from the view of the user, rather than the view of the system. This requires focusing on what is best for the user, rather than what is quickest and easiest to implement.

2.6.4.2 Know the user

This maxim of user centred design appears simple on the surface but in reality is difficult to accomplish. Knowing the user means understanding human behaviour, usually accomplished with observation and application of cognitive and behavioural psychology principles (Hix & Hartson, 1993, p. 29). The designer of an interface needs to understand not just who the user is, but what tasks the user wants to perform, with or without the computer-based system (Hix & Hartson, 1993, p. 30). The design of a system should be tailored to facilitate the use that a user will make of the system – the tasks a user will perform in order to achieve some purpose.

We took several measures in order to better understand the users of the NAL's system. We reviewed the NAL's user survey from 1999, which studied almost 150 researchers. In addition, we created our own survey that was given to a few researchers who have used the Robinson Reports and related documents. We also interviewed some of these researchers, as well as Library staff in order to design a system that we believe meets the needs all parties involved.

2.6.4.3 Involve the user via participatory design

Getting users involved in the interaction development presents tremendous payoffs. From the users' point of view, they will feel more attached to the completed

system and be more likely to use it. In addition, because their comments will be taken into account during the design specification, more of the features they desire will be included. From the designers' point of view, potential users of the system are great assets. Users can help designers understand what tasks the users need to perform, how often the tasks are performed and conditions under which those tasks are currently being performed. In addition, users can tell a designer what they do and don't like about the current system, how they would like to do these tasks differently or can suggest other tasks that they would like to perform (Hix & Hartson 1993, pp. 30-31). We showed our prototype to a number of staff and researchers who gave us input on a variety of issues, ranging from page layout to potential future functionality.

2.6.4.4 Prevent user errors

Anticipating user errors and heading them off is a guideline that can potentially prevent a great deal of user frustration. Greying out menu choices or buttons when they are not available is a good example of this. This is "especially appropriate for direct manipulation, asynchronous interfaces, when many choices exist in the design, but only a subset of those choices might be enabled and available for the user at any given time" (Hix & Hartson, 1993, p. 31). In addition, this is a good example of what is the best for the user not being the easiest design for the programmer.

2.6.4.5 Optimise user operations

Another good guideline is to always offer the user (especially the frequent one) increased efficiency as much as possible. For example, accelerator keys such as "Ctrl-s" can be used as shortcuts for more lengthy commands, in this case using the mouse to select File->Save in MS Word. In addition, more frequent and proficient users may wish to utilize macros to abbreviate their operations. Macros allow users to

define a sequence of frequently used tasks with a single name so that they do not have to type each step of the complete sequence every time they wish to perform the tasks (Hix & Hartson, 1993, p. 32). Along with preventing user errors, optimising operations increases the efficiency and decreases the frustration of the user.

This is an especially pertinent guideline for our project because many researchers are extremely pressed for time and thus need to accomplish tasks as quickly as possible. The worst kind of computer system is the one that takes longer to use than traditional "pen and paper" methods.

2.6.4.6 Keep the locus of control with the user

It is extremely important that the user feels in control of the system at all times, rather than feeling the computer is in charge. The user should have the impression that the computer is prepared to respond whenever he is ready to issue a command (Hix & Hartson, 1993, p. 33). Locus of control has great psychological impact and can strongly influence the user's impression of the system.

2.6.4.7 Be consistent

Consistency is one of the most significant factors affecting usability. Users expect certain aspects of an interface to behave a particular way, and when that does not occur, confusing often ensues. By keeping an interface consistent, users will learn the system quicker and retain more information, which in turn result in more powerful and efficient use of the system. Because the NAL currently has an online card catalogue system that many researchers are using, we made our interface as similar to that one as possible, while still incorporating the new functionality.

2.6.4.8 Keep it simple

The K.I.S.S. (Keep It Simple Stupid) guideline for designing an interface does not live up to its name as it is one of the most difficult to follow. It is not usually

possible to make every command in a system simple to use, but designers should attempt to follow Alan Kay's suggestion of keeping simple tasks simple and making complex tasks possible (Hix & Hartson, 1993, p. 35). As with consistency, keeping an interface simple will allow users to learn and retain the system information quickly and work more efficiently.

When properly implemented, these HCI guidelines help make the system easy to learn and simple to use. They also allow the user to work quickly and commit few errors, which will hopefully result in subjective user satisfaction. In the end, a system with high usability (meaning one that is highly functional yet easy to use) will be the most beneficial for the NAL and users of the system.

2.7 Library Technology

In order to give us an idea of what the future might hold and therefore specify a system for the NAL that would not soon be outdated, we analysed trends in library technology of the past five decades. Fifty years ago, a library's collection consisted of books, journals, newspapers and manuscripts, and the only means of access was the card catalogue. The primary machines found in libraries were typewriters and photostat machines. Anyone who wanted to use a library's information had to go to the library and either use the materials there or borrow them for home use. If a library did not have what the patron wanted, he had two options: request an Inter-library loan (ILL), which could take up to three months, or go to the library that did have the material (De Gennaro, 1987, p. 7). Because libraries were the only source for much of this information, the size of the collection was crucial in determining the value of the library.

2.7.1 The 1950s

In the 1950s, the use of microfilm became widespread in libraries and allowed for the great expansion of a library's resources. Microfilm was a form of publication and reproduction that made materials available to libraries, such as large collections of newspapers and magazines, that otherwise would have been nearly unobtainable (De Gennaro, 1987, p. 7). Before the advent of microfilm, a collection such as the set of New York Times newspapers printed in 1952 would take up dozens of shelves and be highly prone to deterioration and decay. However, the development of smaller, higher quality cameras and processors allowed the same collection to be stored on a few rolls of rugged microfilm and take up a small fraction of the space (De Gennaro, 1987, p. 7). Microfilm allowed libraries to vastly expand their collection without increasing the physical size of the building.

2.7.2 The 1960s

There were two major technological developments that began to affect libraries in the 1960s that came about from the invention of Xerox copy machines. This first was a combination of microfilm and Xerox technology that provided a means for the "economical reproduction and publication of card catalogues by photographing them twenty-one cards to a page and producing them in book form" (De Gennaro, 1987, p.8). A library would periodically produce these books and distribute them to other local libraries. This way, a patron could visit one library and if he did not find what he was looking for, could search through another library's card catalogue on-site. As De Gennaro points out, "the next best thing to having a special collection in one's library was to have a printed book catalogue of that collection" (De Gennaro, 1987, p.8). Another major development in the 1960s was the introduction of the Xerox 914 copy machine. This machine allowed users to quickly, easily, and

cheaply copy library materials and thus use them off-site. In addition, this development meant that scholars no longer had to travel great distances to obtain special manuscripts – rather they could contact the library that had the desired material and have the information copied and sent via mail (De Gennaro, 1987, p. 8).

2.7.3 The 1970s

The 1960s also brought about the advent of computers in the library. At this time, computers were primarily used for electronic card catalogue searches that made vast quantities of up-to-date information available to libraries and their users. However, even though the actual search process was made easier and faster through the use of computers, patrons still had to go to the library to search for materials at that library. It was not until the 1970s that these electronic catalogues were networked so that scholars could search the collections of numerous libraries from one site (De Gennaro, 1987, p. 9). Some of these electronic catalogue systems even began providing e-mail services for ordering books on ILL. As Gennaro writes, "this greatly facilitated and increased the volume of interlibrary loan[s] and made effective resource sharing a practical reality" (De Gennaro, 1987, p. 9).

2.7.4 The 1980s

The application of electronic searching techniques first seen in the electronic card catalogues of the 1970s began to expand to other areas of library materials in the 1980s. Optical disk technology began to enter libraries in the form of CD-ROMS. These media allowed massive collections of information to be placed on a single, durable disk. Libraries began phasing out microfilm and instead put collections of newspapers and journals on these compact disks (De Gennaro, 1987, p. 10). Because the information was now in an electronic format, it was possible to use a computer to search through the contents of the CD. In much the same way that computer

databases replaced traditional card catalogues, CDs replaced microfilm and made possible the “storage, retrieval, communication, and manipulation of vast quantities of research resources in electronic form” (De Gennaro, 1987, p. 10).

2.7.5 The 1990s

The major technological development of the 1990s was not specific to libraries, but rather encompassed the entire world: the internet. While networked computer systems had existed for decades, it wasn't until the proliferation of advanced computer hardware and the development of user-friendly software (such as the Mosaic and Netscape browsers) that the internet really began to take shape (Chen, 1998, p. 6). Because individuals began purchasing their own computers and connecting them to the internet, libraries were able to effectively provide access to their electronic card catalogue systems online. Using the internet, library “patrons” were now able to search for books, journals, etc., containing information they were looking for and possibly request them on ILL, all from the privacy of their own homes.

2.7.6 The Future

While the expansion of card catalogue systems to the internet provided users easier access to searching tools, they were still required to physically visit the library to view or check out these materials. Numerous periodical publishers, such as The New York Times and Washington Post, have recently made their products available online. However, public libraries have not been able to follow suit because most of the information resources they catalogue are copyrighted (De Gennaro, 1987, p. 11). For example, a library cannot put the text of Mark Twain's Tom Sawyer online because that would violate the publisher’s property rights. While some authors are currently exploring the possibility of publishing their materials online, most seem to

have found the potential copyright pitfalls (such as theft or illegal alterations of the text) too great to overcome (De Gennaro, 1987, p. 11). However, the National Art Library is in a unique position, because the British government holds the copyright to, and allows electronic publication of this information (Christopher Marsden, personal communication, 3/15/2001). At the present time technology is such that the resources of a library can now be made available beyond its walls. By transcribing the text of documents into an electronic format, using tags to assign keywords to these documents, implementing a keyword searching system, and then placing this system online, the NAL will be able to provide faster, easier, and more powerful access to their invaluable resources.

3 Methodology

The goal of our project was to design an online search and display system for selected NAL archival documents with embedded linking capabilities. In order to design this system we needed to gain an understanding of the researchers who use these documents. We also needed to study all of the different technologies that could be used to create the system, to ensure that we chose the best options and provided the most functionality for the NAL and its patrons. All of this information helped us to design a software system that increases the functionality of historical documents and makes them more easily available to the scholarly community that uses them.

3.1 Overview of Objectives

The NAL desired a system to display archived material online that would allow text searching and linking to other NAL documents and outside websites. Two very broad steps were needed to reach this result. The first step, to design a system to transcribe the documents into an electronic format, had already been carried out to the NAL's satisfaction by the first IQP team. The second step, to design an online search and display system, was the goal of our team. We not only had to understand the programming aspect required to complete the project, but also the "real world" aspect; we needed to understand who this system was being created for and how they would use it. It was therefore very important for us to gain a good understanding of the mind of the researcher so that the system could be tailored to the intricacies of the research process. If this system could not perform the tasks required by researchers, then it would be of little use to the NAL and its patrons.

It was also very important for us to understand the work done by the previous IQP team. Because the results of our project will be integrally linked with theirs, we needed to understand the tagging system that they developed. It was clear that their

tagging system contained a concise and flexible method of encoding large amounts of information within a document. However, in order to integrate these tags with our search system we had to understand the purpose behind the previous group's work as well as their ideas of what the end result of this system should be.

Defining the objectives necessary to meet our goal was the first and most important step of the work that we did at the NAL. Part of establishing objectives was reviewing our understanding of the project and exploring the expectations of our liaison and the other library staff we worked with. Through telephone conversations, email correspondences, and face-to-face discussions, we were able to garner valuable information that we used to help establish clear and precise objectives that gave our work direction and focus.

Our primary objective was to specify what functions the online library system should provide: what search options should be available, what types of embedded links should exist, and so forth. The recommended program functionality is based on surveys and interviews of actual Library patrons and provides practical information about the researchers' expectations of the online system. The Library will be able to use our findings during the implementation phase of the system to ensure that the desired functions are made available. Our second project objective was to study ways that all of the desired functionality could be provided. We relied on our technical background and research skills to recommend a database table layout and a parser specification that would meet the demands of the users and not strain the budgetary limitations of the Library. Our third objective was to show how this functionality should be presented to the users. We decided to create a Visual Basic prototype of the system that would give the Library an idea of what the final system should look like and how it will function from a user's point of view. In addition to these three main

objectives, we also aimed to determine how the system could be integrated with Z39.50, how XML documents could be displayed directly in a web browser and which documents should be placed online first. This clearly defined list of objectives helped keep our project focused and ensure the NAL greatly benefited from our work in London.

3.2 Gaining an Understanding of the Researcher

In order to design a system that meets the needs of the NAL patrons, we discussed this project with the researchers who would actually benefit from it. We were interested in the level of computer expertise that the researchers possessed, as well as the different uses that they would have for a system of this nature. We accomplished this objective in three ways: reviewing the most recent user survey conducted by the NAL, conducting our own survey of researchers who have used these archival documents, and interviewing some of those researchers.

3.2.1 Reviewing the Most Recent NAL survey

Every so often, the NAL conducts a survey of its patrons. These surveys collect some of the information that we needed to know about the NAL users, and are held over a longer period of time than we would have had to conduct our own. It was therefore logical for us to use the results of their latest survey, which was held from November 30 to December 19, 1999 and a total of 147 responses were received from Library patrons (Appendix A). We analysed this survey for information about the types of research that had been recently done by the public visiting the NAL, as well as any of their problems with the NAL's current cataloguing and organisation methods. We were looking for general trends so that we could learn where people saw shortcomings, some of which could be corrected by appropriate implementation of the online system we were designing. We used this information to help us

understand the research community and to provide some of the background necessary to create a survey specifically for researchers who have used the archival documents that we were working with.

3.2.2 Surveying Users of the Archival Documents

While the NAL has many patrons who routinely use the library to conduct research using their collection of books, periodicals, and other documents, very few of these people work with older archives such as the Robinson Reports. This is due to a number of factors. Many researchers are simply unaware that these archives exist or know that they exist but are not sure where to find them. For those who do know where to find these documents, gaining access to them is still difficult due to the high security, limited hours, and small amount of reading space at the NAL and Blythe House. The often-illegible writing and poor physical condition of these manuscripts further complicate the research process. Finally, there is only one copy of most of these documents so only one person can use them at a time – if this single copy were every destroyed, the information contained within would be lost forever.

Because these older materials are not used as often, there were very few people that we could survey about these documents. We created our own survey with specific questions for researchers who have used the V&A archives at Blythe House (Appendix B). Christopher Marsden, the head archivist of the Museum, was instrumental in distributing our new survey to researchers who had recently worked with the types of documents that the NAL wanted to put into the online system. Most of these people work for the V&A in some capacity and needed to research these documents for Museum business. While we were already aware that we would need to include the ability to search by names, dates, and places, these surveys showed us some of the other features the researchers would like included in the system to aid

them in their work. These surveys built upon our knowledge gained from the NAL user survey and we came up with several ideas and basic concepts that we refined further by interviewing several of the researchers that filled out our survey.

3.2.3 Interviewing Users of the Archival Documents

Interviews with the researchers allowed us to explain our concepts and ideas to those who would be using the online system. They could give responses to our thoughts and we made changes to our basic ideas based on the desires of the system's potential users. Input from the researchers ensured that we fully understood their needs and that we understood the purposes behind the additional features that they would like to be able to use. In this manner, we avoided major miscommunications that would reduce the effectiveness of our online document search and display system. These interviews represented the last step in learning what features the online system needed in order to be most useful to researchers.

All of the information about the research community that we gathered helped us shape the computer system that we designed. While no computer programme will ever replace the mind of the researcher, it can be a valuable timesaving device and research aid. The software, with links to other documents, may in fact be able to show the users another line of thinking that they might have otherwise missed or been unable to pursue. The organisational uses of a computer can often aid those who might overlook a small, seemingly insignificant detail. However, software cannot think, so it cannot form hypotheses and theories in the way that a trained mind can. A computer can perform searches for names and dates and other details to help the researcher find the material that he or she needs. We needed to gain an understanding about these types of searches and how a researcher would use them. Once we had this

understanding, we reviewed the previous IQP group's transcribing system to determine how the tags could be used.

3.3 Exploring the Functionalities of the XML Tags

The previous group created extremely intricate tags in order to accurately describe the information within the Robinson Reports. They designed the tag system to be extensible, meaning that it can be applied to any type of archived documents at the NAL with only slight modifications. We decided to use these tags as keywords for each document because they so accurately and completely described the information contained therein. These keywords will then be placed in a database that will be used to search for documents matching user-specified criteria. The flexibility and detail of this system was integral in designing a comprehensive search system for these transcribed documents. We thoroughly analysed the previous group's tag system in order to gain a full understanding of the uses of their work. We then combined our understanding of the researchers' needs with our knowledge of the tag system in order to determine which tags to include in the database specification process.

We included tags that covered information topics that the researchers we communicated with felt were important. Some tags did not represent information about a subject but rather a method of expanding the written material, such as abbreviation and corrected spelling tags. Part of our analysis was to determine if and how these tags should be represented. This selection was made difficult because of the ability to include "nested" tags, or tags within tags. For example, it is possible to place an abbreviation tag within a person's name tag in order to expand the shortening of "Mister" to "Mr." Choices on how these would be represented were analysed at

this point so that we would know more about our options when we set up the database tables and other programme features.

3.4 Designing the System

By reviewing the data from the NAL survey report and the information gathered from the researchers who have used the archives that we worked with, we developed an understanding of the necessary functionalities the online system needed in order to be useful to the NAL patrons. We then shifted our focus to the computer aspects of our project in order to specify the necessary infrastructure for the system to function as the researchers desired.

Our background research showed us that any online library system contains numerous complicated components. A system such as the one the NAL desires is especially complex because the documents must be searchable by keywords located within the text, not just by subject, author, and title headings as in traditional card catalogue systems such as the NAL already had. There are three major computer components required for a system such as this: a database, a parser, and an interface.

3.4.1 Choosing a Database

Because the database is the core of our system, determining its design and organisation was one of the most important and complex objectives of our project. In order to determine the structure of the database, we created a list of all of the tags to be included in the database tables. Once we had an idea of the structure, we looked at the necessary search capabilities that the database would need in order to read through each of the tags for information. We then finalised the structure of the database, and looked at the various types of available databases to see if any of them would function as desired.

In essence, the database was the foundation for the rest of the project. The complexity of the database enabled the intricacy of the search methods that we planned to use. At the same time, it was important to balance the other issues involved, including cost and future system manageability. The complexity of the tables is an indication of the amount of work that must be done to set up the database, for you not only have to create the tables but initialise the database's internal search protocols to interpret the information. A complex structure is more likely to contain hidden flaws that will not surface immediately, and then be more difficult and expensive to fix when it did appear. However a system that was too simplistic could work perfectly and be easy to create, but be of no use to anyone due to limited functionality.

3.4.2 Choosing a Parser

Once we decided on the details of the database, the next step was to choose a parser for the documents. Because the parser needs to be able to read through the transcribed documents, extract the tags and place the information located within the tags into the appropriate fields in the database, the parser that we needed had to be compatible with both the documents and the database. There were several types of existing parsers, some free and some proprietary that we looked into using, and we also examined the benefits and drawbacks of trying to custom-design a parser for this system. After we explored all of the available options, we chose the parser option that best met our criteria.

Choosing the parser was the easiest step of our work, but still important. In many cases, a parser was unusable because it would not interface with the database and the XML, or because it was too simplistic to sort the data as we required. Other

parsers however, would function in the manner desired, and we then had to analyse each in detail to be sure that we chose the best.

3.4.3 Designing the Interface

Designing the interface was the last piece of the system design work as it was necessarily influenced by the previous decisions made about the database. The interface was basically a cover placed over the system to let people easily feed information to the database's search module. The interface has to allow for each type of searching that we wanted to include within the system, and also display all of these results in a format that the users could quickly read and understand. In order to design these interface components, we studied the principles of human-computer interaction. A study of these sources helped us shape an easily understandable layout for the separate pieces of the interface.

Another part of the interface that we designed was the collection of protocols of the online display of the documents. We again needed to balance the different functionalities that we planned to include with ease of use and manageability. Two different tools that we included were embedded linking capabilities and quick display boxes to show a small amount of basic information about certain keywords. However, we needed to be sure that such options did not clutter up the screen and become visually distracting for a person reading the documents.

In a way, the interface could be considered the most important component, because if the interface is unusable, then nobody would take advantage of this software, and so the NAL would not be providing the additional access to documents that they would like to. In reality though, all of our work was of equal importance, because in the end, each part, the database, the parser, and the interface, had a role to play in the system. A good interface would mean nothing if the system behind it were

not of use to people, and the most comprehensive search system in existence would never be accessed if nobody could figure out how to use it.

3.4.4 Creating a Visual Basic Prototype

Once we established all of the parameters for the entire online system, we worked to create a Visual Basic prototype of the interface. This prototype demonstrates how a person would use the system to search through the online documents, and the different display methods that we designed. This allowed the NAL to see our thoughts on exactly how the online system would function and a chance to suggest changes that would make the interface more useful. After the review of our prototype, we made changes based on the suggestions and comments we received.

3.4.5 Conclusion

Our results represent the design of the online system that we believe suits the NAL and its patrons. Our study of the researchers and their needs contributed to a design for an online resource that will be extremely useful in the research environment. The prototype and the programme specifications show the integration of today's technology with the historical documents and their link to the past.

4 Results & Analysis

This chapter presents the results of our work at the National Art Library. It includes the findings from our analysis of the users of the NAL and Blythe House as well as the analysis of the TEI and EAD tag sets. We used the information from these analyses to design the online system, the specification of which is found in the Conclusions chapter.

4.1 Analysis of NAL Patrons

Understanding the researchers who use the NAL was integral in specifying a successful design for the online system. Consequently, we gathered relevant information from numerous, varied sources: the 1999 NAL reader survey, our own survey of select researchers, and interviews with some of these researchers. The NAL reader survey provided us with general data about the NAL and its patrons, while our own survey and follow-up interviews brought out more detailed information about the needs of archive researchers. These three sources of information provided us with a better understanding of what types of system functionalities the researchers desired, which aided us in our design specifications.

4.1.1 Importance of the NAL

The NAL is a heavily used library and is growing in popularity – according to the Library's reader survey 27% of patrons visited the library more than 10 times in the past year and 37% of patrons planned to visit the library more than 10 times in the coming year. The items in the NAL are extremely important to many researchers as indicated by the 55% of readers who either agreed or strongly agreed with the statement "In my subject the NAL collection is unparalleled." In addition, 73% readers agreed or strongly agreed that access to the NAL is essential to their research

and 67% agreed or strongly agreed that the material available in the NAL is unavailable elsewhere. Clearly, the use of NAL documents is crucial for many researchers and an online text display system will help make these invaluable documents far more accessible to the research community.

4.1.2 Limitations of Current Access

Although documents housed at the NAL are very important to many researchers, getting to the Library is difficult for many of them. 42 readers, representing 28% of the total polled in 1999, listed their "normal residence" as outside the city of London and over half of this group live outside of the UK. However, even if a researcher is able to make the trip to the NAL, he is presented with other obstacles preventing easy access to historic documents. On a scale of 1 (poor) to 4 (good), 22% of readers polled by the NAL ranked the waiting time for book delivery 2 or lower and 17% also ranked the time spent queuing (in line) a 2 or lower. The other major complaint was the opening hours of the Library, which were ranked as 2 or lower by 38% or readers surveyed. We believe that providing the text of documents online may save commuters precious time and energy and that an online text system with fast, concurrent, 24 hour-a-day access to NAL documents would be extremely useful to many researchers.

4.1.3 Other Sources for Research

According to the 1999 reader survey, 71% of all readers are spending at least some of their time searching for material that is unknown to them – they don't know where or even if the information exists. Partly due to this, many Library patrons also use other libraries for their research, the most popular being various university libraries (65%), public libraries (46%) and the British Library (46%). For this reason, we would like the NAL's online text system to be compatible with similar systems

throughout the world, as has already been done with the Library's online card catalogue. This integrated searching feature can be accomplished by following CIMI protocols such as Dublin Core and Z39.50, both of which provide the ability to search a worldwide academic community of resources simultaneously. Providing online access to selected archives with comprehensive searching capabilities can drastically reduce the time readers spend at the NAL looking for information that may or may not exist and integrating the system with CIMI protocols will help to connect the Library with other sources of information over the internet.

4.1.4 Researchers' Computer Abilities

While providing a high degree of functionality in an online library system can be very helpful, it does no good if the intended users of the system are not comfortable with computers and the internet. According to the 1999 survey, most users of the Library are at least somewhat computer literate as indicated by the 75% who used the NAL's computer catalogue and 1% who used Library databases on CD-ROM. In addition, 17% of readers surveyed used the NAL's WWW pages and 14% used its email enquiry service. Based on this data, as well as the researcher activity we have witnessed in the past seven weeks while working in the Library's Centre Reading Room, we are confident that researchers are comfortable with computer technology and will use the online text system extensively.

4.1.5 Researchers' Desires for an Online Text System

According to the NAL reader survey, 24% of readers were using NAL resources to study the Museum itself. In addition, 15 of the readers surveyed indicated that they had used the V&A Archives at some point. This indicates strong interest from at least a few researchers in the documents that have already been transcribed and would initially be placed online such as the Robinson Reports and Art

Referee's Reports. What follows is an analysis of these researchers' desires for the online library system.

4.2 Analysis of Archive Researchers

Based on the 1999 NAL reader survey, it would seem that an online text display system would be highly beneficial to the NAL and its patrons. Unfortunately, all of the text from all of the materials in the Library cannot be transcribed and placed online immediately and the NAL must select a few manuscripts to use as a starting point. These will not only be immediately useful to researchers, but also be used as examples of how to place other documents online in the future. The Library chose four sets of documents, three of which are housed at the V&A archives at Blythe House, to transcribe and place online first: the Robinson Reports (1863-1868), Art Referee's Reports (1868-1886), Abstracts of V&A correspondences (1864-1914) and Board Minutes of the Science & Art Department (1852-1892). In order to understand the users of these particular documents, and therefore design an appropriate system for them, we created a survey asking archive researchers what they would find most useful in the online text system and later interviewed them to garner more specific information on their capabilities and desires. Our survey, along with a tabulated list of results, can be found in Appendix B.

Because the people we interviewed are experts in their field, they were able to recommend many online sources (both inside and outside the Museum) that would be useful to link to the library system. The V&A has created a Collections Information System that consists of an electronic inventory of almost all Museum objects, excluding Library collections and items from the prints and drawings archive. The Museum is currently working on placing this system online, and the staff we interviewed thought this would be an extremely useful source to link into the online

library resource we designed. The Museum is also working on an electronic picture archive, which will contain digitised photographs of Museum objects. Linking the online library to this system would provide users with pictures of the actual art objects they are reading about. These internal links, combined with the links to the online library catalogue and NAL website included in our prototype, greatly contribute to the power of the system we designed by increasing the number of portals to relevant information.

There are also a number of online resources outside of the V&A that staff thought would be useful to researchers. Standardized research tools such as a Thesaurus and Dictionary would be useful to researchers unfamiliar with antiquated terminology often found in archived documents from the 19th century. Specialized Art Dictionaries, such as Grave's Dictionary of Art, would be useful for more specific terms used to refer to art objects. Furthermore, online biographies of people mentioned in the online texts would also be useful for those searching for information on a particular person. Finally, it was also recommended that the system contain links to company sources such as auction houses where Museum objects were purchased. While the final specification of our system does not specifically include these links, the design is incredibly flexible and can easily accommodate these links if the Library decides to add them in the future.

4.3 Analysis of TEI Tag Set

In order to design an effective online library resource for the research community at the NAL, we needed to be sure that we used all of the tools provided by the electronic transcriptions, i.e. the tags, to their fullest potential. Therefore, we performed a detailed analysis of the two tag sets that the documents were marked up with, the TEI tag set and the EAD tag set (Appendix C). Because there are

discrepancies between the two tags sets, each set needed a different table design for the database in order to incorporate all of the information that the tags contained. The TEI set, designed by the previous IQP team for use with the full text of documents, was the first set that we analysed.

The main purpose of the contextual tags of the TEI tag set is to mark the key words and phrases within the text, which allows a computer to recognise them. These tags provide the main method of implementing a search system for document text by marking various names, dates, and other important pieces of information within the text itself.

4.3.1 Names

The largest group of contextual tags are the separate types of <name> tags. There is a basic <name> tag that can be used to mark a name of any type. However, the TEI standard allows for the modification of the <name> tag in order to indicate a specific type of name. In the tag set defined by the last group, there are seven different types of <name> tags, which refer to separate types of names: collection name (<collectionName>), event name (<eventName>), material name (<materialName>), object name (<objectName>), organisation name (<orgName>), person name (<persName>), place name (<placeName>). There is another type of <name> tag that was specially created by the previous IQP team to indicate places that are not geographical locations or cities, called the specific site name, represented by the tag <name type = site>. This tag is used to indicate sites such as telegraph stations and train stations, and is actually just a generic <name> tag that has the "type" attribute set to site. Each of these eight types of <name> tags are formatted the same way using the <name> standard of TEI.

Within all of the <name> tags there are two attributes that can be added to the tags. The “reg” attribute indicates the regularised version of the name. This option allows an abbreviated name to have a regularised version within the tag. The “type” option is the other modification to the <name> tags. In the <name type = site> tag, the “type” attribute has been permanently set to “site” to specialise the tag. Except for this specialisation, the “type” attribute is used elsewhere only with the <persName> tags, which indicate a person’s name. The “type” attribute defines the relationship of the person to the document; such as author, subject, reader, or recipient. This attribute can be used to specify any relationship between a person and the document.

4.3.2 Dates

There are two types of TEI tags that deal with chronological references; the <date> tag and the <dateRange> tag. The <date> tag is used only with a single date, while the <dateRange> tag indicates a span of time. The date tag represents specific dates such as “April 14th, 1863,” or “June 1875.” The <dateRange> tag is used for time ranges such as “1871-1892”, or expressions like “the 15th century.” These tags are used for every date or date range that appears within the transcribed text. There is no special date tag that separates the date that a document was created from any other date mentioned within the document. Instead, placing the <date> tag within an <opener> tag indicates the date that the document was written. Other than this change, the date of creation functions as a normal <date> tag in every way.

The “value” attribute is always included with the <date> tag to standardise the way the date appears. No matter how the date was written in the text, within the tag itself the date will be specified as Year-Month-Day (YYYY-MM-DD). If the full date is not written, then the “value” attribute will be written as YYYY-MM if the month and year are given, or YYYY if only the year is given.

The <date> tag also contains the "certainty" attribute, an indication of how precise the transcriber feels the date is. The "certainty" attribute is given a value of 0 to 100, 0 indicating that the transcriber is sure that the date is wrong and 100 indicating that the transcriber is sure that the date is correct. This attribute is only used if the transcriber wishes to indicate his or her opinion about the date.

The "from" and "to" attributes are used with the <dateRange> tags to standardise the dates. The "from" attribute indicates the starting date of the range, and the "to" attribute gives the ending date. The dates are given in the same form as the "value" attribute, YYYY-MM-DD, if the dates are given that precisely or YYYY-MM or YYYY if that is all the information listed.

The "exact" attribute is the transcriber's opinion about how accurate the time span is. This is not a number value, as with certainty, just a yes or no answer. Either the "from" value is exact, the "to" value is exact, neither are, or both are. This is always used within the <dateRange> tag. Often, the transcriber is forced to use his or her own opinions at to what the date range is. For example, if "circa 1860" appears within the text, then the transcriber can indicate that this represents a range from 1850-1870, but this is inexact, because the transcriber is assigning the range. However if "1863-1867" appears within the text, then the transcriber can list this range as exact because it is given directly by the document.

4.3.3 Cost

The <cost> tag is used to indicate the price of an art object, either the price actually paid or the price that was requested by the seller. The <cost> tag is only used for prices of art objects, no other monetary expenses. The <cost> tag itself contains no reference to the art object, just the numerical value of the price. The <cost> tag has no attributes.

4.3.4 Registered Paper Number

The Museum has a method of labelling archived material by stamping each page with an RP number, or Registered Paper number. This number is an indication of the chronological order in which the papers were filed into the archives, although this does not always follow the chronological order in which they were written. The RP number is represented with the <biblScope type = superceded>. The <biblScope> tag is the bibliographical information tag, and the "type" attribute is set at "superceded" to differentiate the RP number from the other forms of bibliographical indexing that are available with the TEI tag system.

4.3.5 Museum Object number

The <objectIdentifier> tag is used to mark the Museum Object number. This number is the Museum's method of cataloguing all of their items and art objects. Every item has a unique Object number, which is used in the Collections Information System, a computerised resource for Museum staff to look up information about the various objects owned by the V&A. The Museum Object number is useful because it provides a link between items mentioned within historical documents to current V&A information about those objects. The <objectIdentifier> tag contains the "reg" attribute to standardise the Museum Object numbers, which have been written in different formats over the years.

4.3.6 Document Title

The <title> tag denotes the title of the document, whether given by the author or at a later date. The title of a document is the most basic piece of information about the document, divulging the subject of the work and perhaps even a little of the purpose of its creation. The <title> tag has no attributes.

4.3.7 Abbreviations

Abbreviations within the documents are marked with the <abbr> tag. All abbreviations with the documents can be marked with this tag. The <abbr> tag contains the "expan" attribute, which will contain the expansion of the abbreviation. There is also the "type" attribute, which indicates the type of abbreviation, such as a contraction, a title abbreviation, or an acronym. Finally, there is the "cert" attribute, used to indicate the transcriber's certainty that the full form of the contraction is correct. The "cert" attribute is based on a range of 0 to 100, with 100 indicating that the expansion is definitely correct. The "cert" value is usually 100, except in some instances where a contraction could indicate several words and the transcriber has worked from context, or in cases where an antiquated term has been used, and the transcriber simply isn't sure if the expansion is correct.

4.3.8 Antiquated Text and Spelling Mistakes

The documents that have been transcribed often contain antiquated or irregular text, and often some spelling mistakes. In cases where the spelling of a word has changed over time, the <orig> tag is used to indicate antiquated or irregular text with the documents. The <orig> tag contains the "reg" attribute, which contains the modern spelling of the antiquated word or phrase. Similarly the <sic> tag is used to mark words that the writer of the document misspelled. The <sic> tag contains the "corr" attribute, which will contain the correct spelling of the word, and the "cert" attribute, which, on a scale of 0 to 100, indicates the transcriber's certainty that his or her spelling is correct.

4.3.9 Damaged and Supplied Text

There are several tags that deal with illegible or unreadable spots within the documents. Some of the tags indicate damage to the text, due to age or other causes,

inkblots, sections where text has been crossed out, or places where the text is missing (the paper has torn). These tags represent physical facets of the documents, and are not contextual, but there is a companion tag, the <supplied> tag, indicating that the transcriber has added his or her own text, attempting to guess by context what text had been lost. The <supplied> tag has a "reason" attribute, to indicate the cause of the damage.

4.3.10 Nested Tags

Several of the contextual TEI tags can be nested within each other. Some tags, such as the various types of name tags, the date tags, and the cost tag are discrete and will stand alone and not be nested within another contextual tag. The abbreviation tag, the antiquated text tag, the spelling mistake tag, and the supplied text tag can all be nested within other tags. For example, an abbreviation tag can be nested within one of the types of name tags to indicate that the name contains an abbreviation. The tags that can be nested are those that do not indicate a specific type of data, but rather editorial tags that supply additional information about the text.

4.4 Analysis of EAD Tag Set

The EAD tag set is a much more complex tag set than the TEI standard that was used by the previous IQP group. The main difference is that the TEI set that we worked with was narrowed down for a specialised use by the previous group, and so contains tags designed to deal with the Robinson Reports and similar art history documents, while the EAD tag set had not been refined to these specific document types.

The EAD standard contains a large group of tags to deal specifically with archival abstracts. These tags are used to note the physical history or ownership of the archives. The formatting tags are more complex, because the abstracts can be in

many formats, such as table layouts or paragraph form. There are also a number of tags that concern the history and makeup of the archive itself rather than the information contained in the archive. However, there are still contextual tags that can be used to mark important words within the text, so that additional information can be stored as metadata, to make the archives more useful.

The main focus of our EAD tag set analysis was to identify the contextual tags that identify information within the documents, the tags that are used within the text to mark the key words and phrases. Just like the TEI tag set, these tags are used to mark names, dates, and other important pieces of information, and can contain additional metadata that can be used by a database to more fully interpret the documents.

4.4.1 Names

As with the TEI tag set, the largest group of contextual tags within the EAD tag set are the different naming tags. There is the basic <name> tag, used when the transcriber is unsure of what type of name is being mentioned by the document. There are also several specialised forms of the <name> tag, each with a specific purpose. The <corpname> tag refers to the name of a company or corporation. The <famname> tag is used to indicate the name of a family or lineage. The <geogname> tag indicates any geographical name, city, country, a specific building or monument, or a natural feature of some type. The <persname> tag is used for the name of a single person.

All of the different types of name tags have the same two contextual attributes. The first is the “normal” attribute, an optional attribute that the transcriber can use to fill in the standardised form of the tagged name. This is especially useful in the abstract archives because names are often abbreviated or shortened within the text.

There is also the “role” attribute, which indicates the relation between the name and the document, or the relationship between the name and the materials being described by the archive. For example, the “role” attribute could either note that someone is the subject of the document, or that the person took the photographs that the document is describing.

The <title> tag is similar to the <name> tag, but is used specifically to mark the title of a work being described in the text. In this context, the title refers to the formal name of whatever object, such as a book or a painting, is being described within the document. The <title> tag has only the “normal” attribute, to supply a standardised form of an object’s name.

4.4.2 Dates

There are two <date> tags within the EAD tag set, <date> and <unitdate>. The <unitdate> tag is used to mark the date that the archive was created, while the <date> tag is used for all other dates within the document. These date tags can be used to specify both a single date and a span of time. They contain three contextual attributes: “normal,” “type,” and “certainty.” The “normal” attribute is used to specify a standardised form of the tagged date, either in YYYYMMDD for a single date or YYYYMMDD – YYYYMMDD for a range of dates. For the <date> tag, the “type” attribute indicates a more specific reason for the date, such as a publication date or a birth or death. For the <unitdate> tag, the “type” attribute indicates whether the dates are inclusive of the archive creation, or simply the predominant dates. For the <date> tag, the “certainty” attribute indicates the accuracy of the date, given by a word to indicate if the date is not meant to be precise, such as “circa” or “approximately,” rather than the 0 – 100 scale of the TEI tag set. For the <unitdate>

tag, the “certainty” attribute is used to specify if the date is given by the document or if it is an estimation by the archivist.

4.4.3 Registered Paper Number

The EAD tag set contains a specialised tag defined by the V&A staff that use it to tag documents to indicate the Registered Paper (RP) number on the transcribed abstracts. The tag appears as <unitid type = previous>. The <unitid> tag is used in the EAD tag set to indicate any set of characters that represent a unique document reference, such as the RP number. The “type” attribute is set to previous in order to separate the RP number from the other forms of document identification that are currently used by the V&A.

4.4.4 Document Title

The title of the transcribed documents is marked with the <unittitle> tag. The title of the document can either be given by the original author or supplied at a later time. The title of a document represents, at a very basic level, the subject and sometimes the purpose of that document. The <unittitle> tag has no attributes.

4.4.5 Abbreviation and Expansion

There are two editorial tags in the EAD tag set to deal with abbreviations. The <abbr> tag is put around an abbreviation, such as an acronym or contraction, or a shorthand abbreviation, and then the “expan” attribute is used to give the expanded form. Similarly, the <expan> tag is used when the expanded form of a commonly abbreviated word or phrase, such as an acronym, appears written out in its full form. Then the “abbr” attribute is used to give the abbreviated form.

4.4.6 Nested Tags

Unlike the TEI standard, these EAD contextual tags are not often nested within one another. It is more common to use the “normal” attributes of the tags to

hold information. The abbreviation and expansion tags, for example, are almost never nested within the different name tags. The “normal” attribute is used almost exclusively to handle any abbreviations that occur within the tags. However, there are some instances where the tags could be nested or the “normal” attribute could be used, depending on the desires of the transcriber.

5 Conclusions & Recommendations: System Design

We used the information that we gathered from the researchers and our analysis of the tag sets to design the specification of an online library system for the NAL to implement at a later date. Our design is generic to all commercial database systems, such as Oracle and Sybase, so the NAL can implement this search system with a database of their choosing.

5.1 Search Field Parameters

We decided to make two types of search pages, the simple search and the advanced search. These are the two types of search protocols found on almost every search engine in existence. Each provides a useful method of narrowing the electronic texts to find the information that a particular researcher is interested in.

5.1.1 The Simple Search

The simple search is a powerful tool for a researcher using the system because it searches through all of the database tables for whatever words or phrases that a researcher requests. The researcher cannot narrow down the search fields, so running a simple search will return every occurrence of the searched word or phrase. This search is very basic, often used as the first step when conducting online research. After a simple search is run, and a researcher can discover the general trends in documents that deal with his or her topic, he or she may move on to the advanced search to narrow down the search parameters and move towards a smaller group of documents that focus more closely on the topic.

Because the <cost> tag contains no attributes, it is a very simple tag with limited use. We have decided that the database will not search it on a basic search, but rather only search it when requested for an advanced search. The <cost> tag is very limited, so during a basic search it would call up too many results. The cost

search field can either be filled in as a single value or as a range of values, in case the researcher only knows an approximate price. Similar to the problem of ranges within a date search, a similar type of system will have to be set up for the cost search, to interpret a range and find all of the appropriate values.

5.1.2 Advanced Search Fields and Tag Correlations

The advanced search page provides the researcher with a number of search fields to narrow down the range of his or her search. It is also be possible to run a search for multiple options at the same time to find documents that meet several separate criteria.

The search fields are:

Name of a Person
Name of a Place
Name of an Art Object
Name of an Organisation
Name of a Museum Collection
Material used to create Art Object
Name of an Event
Cost of an Art Object (Exact or Range)
Date (Exact or Range)
Registered Paper Number
Museum Object Number

The search parameters each correspond to at least one of the TEI or EAD tags. However, some search fields have a tag in the TEI tag set but not in the EAD tag set, or vice versa. The search engine will have to be set up so that it knows which of the tables to search when a specific search is requested.

5.1.2.1 Name of a Person

The "Name of a Person" search runs a search through the TEI <persName> tag database table, the EAD <persname> tag database table, and the EAD <famname> tag. These are the only tags that refer to a personal name.

5.1.2.2 Name of a Place

The “Name of a Place” search runs a search through the TEI <placeName> tag, the TEI <name type = site> tag, and the EAD <geogname> tag. The TEI place name (<placeName>) tag refers to the name of a city or country, or in some instances geographic features such as mountains or rivers. The specific site name (<name type = site>) refers to the name of specific buildings or monuments. This could be a small grammatical distinction for a researcher, and might easily cause a lot of misunderstanding and possibly result in a search missing a vital document. To eliminate this potential problem we have decided that these tags will be separated within the database, but the search engine will know to search of both fields. The EAD <geogname> tag covers all names that refer to a place or location.

5.1.2.3 Name of an Art Object

The “Name of an Art Object” search runs a search through the TEI <objName> tag and the EAD <title> tag. These are the only tags that refer to the names of art objects.

5.1.2.4 Name of an Organisation

The “Name of an Organisation” search runs a search through the TEI <orgName> tag and the EAD <corpname> tag. These are the only tags that refer to organisation, company, or corporation names.

5.1.2.5 Name of a Museum Collection

The “Name of a Museum Collection” search will search through the TEI <collectionName> tag. There is no corresponding EAD tag for a collection name.

5.1.2.6 Material used to create Art Object

The “Material used to create Art Object” search will search though the TEI <materialName> tag. There is no corresponding EAD tag for the material used to create an art object.

5.1.2.7 Name of an Event

The “Name of an Event” search will run a search through the TEI <eventName> tag. There is no corresponding EAD tag for event names.

5.1.2.8 Cost of an Art Object

The “Cost of an Art Object” search will run a search through the TEI <cost> tag. There is no corresponding EAD tag for objects’ costs.

5.1.2.9 Date

The “Date” search will run a search though the TEI <date> tag, the TEI <dateRange> tag, and the EAD <date> tag, and the EAD <unitdate> tag. These are the four tags that deal with dates and chronological references.

5.1.2.10 Registered Paper Number

The “Registered Paper Number” search will run a search through the TEI <biblScope type = superceded> tag and the EAD <unitid type = previous> tag.

5.1.2.11 Museum Object Number

The “Museum Object Number” search will run a search through the TEI <objectIdentifier> tag. The Museum staff working with the EAD tag set plans to specify an EAD tag for Museum Object number, but they have not yet decided which tag to use.

5.2 Database Table Design for Text Searching

Only the tags that deal with names, dates, cost, the Registered Paper Number, and the Museum Object number will have their own tables within the database

(Appendix D). These are the tags that give definition to the information that they contain. The abbreviation tag, the antiquated text tag, the spelling mistake tag, and the supplied text tag are editorial tags that give no information about the text on their own. Their purpose becomes apparent when they are nested within other tags, providing more detail about the main tag and enabling a more comprehensive search. These nested editorial tags appear as subtables within the main tables of the informational tags.

For the name tags, the TEI "reg" and the EAD "normal" attributes are extremely useful. By telling the database to search through the attributes, we can ensure that a name is not missed because it was shortened or only initials were given within the document.

For the various date tags, the standardised forms of the dates are very important. It provides the database with an easy way to search all of the dates without worrying about the document text. The database search engine will need to be told how to search through the date tags to interpolate a range of values and the database will have to be programmed to interpret given values to find all of the documents that match.

The tables can be designed to grow dynamically, so that a new subtable can be automatically created for each new nested tag. The programmer will not have to add each nested tag manually to a table, the tables can grow and change without human intervention as more documents are transcribed and added to the system. Each type of tag has a format attached to it, called an object (Appendix D), so that once a tag is recognised, the format will be applied to the database table structure, and then the information held by the tag will be placed in the appropriate area.

5.3 Parser Design

There are no commercially available parsers that will function in the manner required to set up this database. The parser must be able to interface with both the XML documents and the database that the NAL uses to set up this search system. Creating a parser is not a complicated task, and once the NAL has chosen a database, a programmer can quickly create the parser to fill in the database tables. Because the database will be able to grow dynamically, the parser must be able to tell the database to create the appropriate subtables as it scans the documents and puts the information into the main tables.

5.4 Interface Design

The goal of the interface, as with any interface, is to provide the user with a high degree of usability, which is a combination of five user-oriented characteristics: ease of learning, high speed of user task performance, low user error rate, user retention over time and subjective user satisfaction. We accomplished this goal of high usability by following the HCI guidelines discussed in section 2.6: practicing user-centred design, knowing the user, involving the user via participatory design, preventing user errors, optimising user operations, keeping locus of control with the user, being consistent and keeping operations simple.

The first thing one will notice about our interface is how similar it looks to the current online library catalogue. We used the same colour scheme and page layout because many users of the NAL are already familiar with that system. This consistency will make the online text system easier to learn and use, as well as help prevent user errors. In addition, we believe that this system would best be used as a complement to the current online catalogue and NAL website, rather than a stand-

alone resource and for this reason have incorporated links to both the catalogue and website in the online text system.

There are two parts of the system that users can use to search for documents, the new search (or simple search) page and the advanced search page. The simple search page contains two data fields for the user to fill in, the archive he would like to search and the word or phrase he is looking for within that archive. Initially only the Robinson Reports, Art Referees Reports, abstracts of V&A correspondences and board minutes from the Science & Art Department will be available online, but the interface is easy to expand to accommodate future transcriptions.

The advanced search page is an important element of the interface as it allows users already familiar with the documents to refine their search and view more appropriate results. Like the simple search page, the advanced search page first asks the user to select an archive to examine. However, the main power of the advanced search is that it allows the user to select indexes to search. Based on our surveys and interviews we determined that the indexes the users would find most useful are the registered paper number, cost, date and date range, museum object number and several name tags (art object, collection, event, organisation, material, person and place names). As with the archives, the interface is easy to expand if the Library decides other indexes would be useful. The other major power of the advanced search page is that it allows the user to refine his search with the Boolean operators “and”, “or” and “not.” We believe that the combination of the simple and advanced search pages will increase usability by allowing both novice and expert users to quickly and easily find the information they are looking for.

After entering the search criteria and pressing the “search” button, the computer tabulates the results and returns them to the user. The search results page

contains a number of important features that make the online library resource more user-friendly and functional. The word or phrase that the user most recently searched for is displayed at the top of the page because it provides a point of quick reference without having to hit the "back" button. Below this line is the area where the documents matching the user's search criteria are displayed. Based on suggestions received during our interview with Christopher Marsden, the title of the document is in bold, followed by the date of the original document's creation and its registered paper number. Mr. Marsden also thought it would be useful for the researcher if we included a brief summary of the document under its search results listing. Unfortunately, electronic summaries of the documents do not exist so this was not possible; instead, we placed the first few lines of the document text under its search results heading to give the user an idea of what type of information the document contains. Another useful feature found on the search results page is the pull-down menu at the bottom that allows the user to conduct the same search on google.com or yahoo.com. This menu also allows the user to perform a search on the online versions of the V&A Collection Information System or Photo Archive, both of which are currently under construction. As with all of our design, this menu is easy to extend to provide links to other resources in the future. These links provide yet another portal for the researcher to garner information and contribute to the power of this online library resource.

The part of the interface that displays the actual document is arguably the most important because it is the major source of information for the researcher and the main reason why he is using the system in the first place. In his survey, Mr, Marsden recommended that this page contain a running header with the title, date and RP number for the document being displayed. Below this header is the area where the

text is displayed, and the default view shows the keyword links in blue text. While not functional in our prototype, when clicked on these links will perform a search for the tagged words. If the user just wants to read the text of the document and feels that the blue keywords are distracting he can turn them off by selecting the "hide links" button at the bottom of the page. At the opposite end of the spectrum, if the user is *only* interested in the tagged keywords he can select the "show tabulated display" option to display the keywords categorically. Our main goal with the text display page was to allow users to customize the view to their liking, which should result in higher usability through subjective satisfaction.

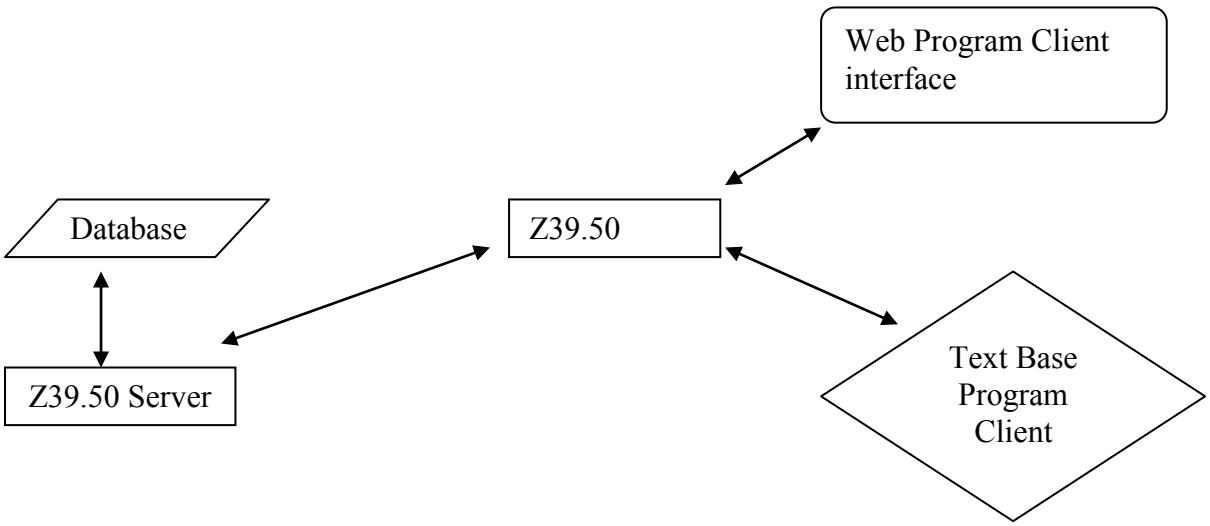
5.5 Accessibility Issues

Besides simply designing this search system, we researched the accessibility issues that concern the NAL. There are several organisations around the world that are attempting to create unified search systems that provide easy access to multiple electronic archives. These groups have all taken different approaches to the problem, and have done interesting work. The two main ones that we looked at were Z39.50 and Dublin Core.

5.5.1 Z39.50

Z39.50 is a standardised protocol used to search and retrieve information from a database. It allows other institutions to interact with the database by using a Z39.50 compliant client. This is a standard way of searching for one type of information on multiple types of databases, instead of searching on one database at a time. This protocol also allows different ways of interfacing with it because it is a standard way of communicating. Rather than requiring multiple programs for communication to a database, you make multiple programs communicate to the Z39.50 client, and then the

Z39.50 client returns information from the database to the program client. The diagram below shows what other interfaces can be made with the Z39.50 protocol.



The diagram above only shows one database, but there can be more than one server connection to a client.

For example, the Z39.50 server can be told to find a person's name, then when it searches each database, it will look through and find the name it needs by translating the query from Z39.50 protocol to the database's language. Then the database will return the result to Z39.50 server, and the server will translate it back from the database language to it's own and return the result to the user.

The web programs can also learn how to do other types of searches, and still use one type of protocol, to communicate to the other databases that exist out there in the world. The client just needs to know where the other Z39.50 server is and then it can link to it and query for data and search into the database that exist with the Z39.50 server.

In order for the Z39.50 server to work with our database it must understand how to talk to an SQL database. There is Z39.50 software that exists to communicate

with an SQL database. However, the NAL will need to create an in-house implementation of the protocol to interact with the database. The database tables for the types of reports we worked are complex due to the amount of information contained by the tags, and it is necessary to return simplified results back to the Z39.50 server.

5.5.2 Dublin Core

Dublin Core is a separate attempt to create a worldwide search protocol, by creating a standard set of simple tags. Dublin Core is a way of tagging documents using a fewer number of elements, basically simplifying the method to represent metadata. It is a standard used by museums so they can catalogue what they have in their institutions. Because there are a fewer number of tags and tag attributes in the Dublin Core standard, it is easier to tag a large group of documents quickly. However, for the NAL's purposes, Dublin Core is too simplistic. Much of the functionality that comes with the TEI and EAD tagging systems would be lost with the implementation of the Dublin Core system. The Dublin Core tag set is much more limited than the current tagging sets that have been used to mark up the NAL's electronic documents.

5.6 Continued Expansion of the Online Library

All of these design specifications are meant to meet the requirements and desires of the researchers and potential users of this system while still meeting the NAL's criteria of simplicity and easy implementation, in effect balancing functionality with manageability. Our efforts have resulted in a system that is extensible and scalable so that as the NAL's collection of electronically transcribed documents grows, the online library can grow with it.

6 Bibliography

Baker, M., & Richardson, B. (1997). A Grand Design – The Art of the Victoria and Albert Museum. New York: Harry N. Abrams.

Banks, F. R. (1973). The Penguin Guide to London. Baltimore: Penguin Books Ltd.

Bosak, J. (1997). XML, Java, and the Future of the Web. [Online]. Available: http://www.ibiblio.org/pub/sun-info/standards/xml/why/xmlapps.html [2001 February 11].

Burnard, L. (1995). TEI Lite: An Introduction to Text Encoding for Interchanging [Online]. Available: http://www.uic.edu/orgs/tei/lite/index.html [2001, January 30].

Burton, A. (1999). Vision & Accident – The Story of the Victoria and Albert Museum. London: V&A Publications.

Chen, Su-Shing. (1998). Digital Libraries – The Life Cycle of Information. Columbia: Better Earth Publisher.

CIMI [Online]. Available: http://www.cimi.org [2001 January 30].

Davis, D. (2000, September 24). The Virtual Museum, Imperfect but Promising. The New York Times, Section 2, 1.

De Gennaro, R. (1987). Libraries, Technology, and the Information Marketplace. Boston: G. K. Hall & Co.

Dodds, D. (2001). (personal communication by telephone, January 2001).

Electronic Text Center at the University of Virginia. (2000). TEI Guidelines for Electronic Text Encoding and Interchange (P3). [Online]. Available: http://etext.virginia.edu/TEI.html [2001 January 30].

Fidelman, M. (1997). All-out Internet Access – The Cambridge Public Library Model. Chicago: American Library Association.

Fletcher, P., & Bertot, J. (Eds.). (2000). World Libraries on the Information Superhighway: Preparing for the Challenges of the New Millennium. Hershey: Idea Group Publishing.

Fontanella, L. (2001). (face-to-face interviews, January & February 2001).

Furrie, B. (2000). Understanding MARC Bibliographic: Machine-Readable Cataloging [Online]. Available: http://lcweb.loc.gov/marc/umb [2001 January 30].

Gracy, D. B. (1977). Basic Manual Series – Archives & Manuscripts: Arrangement & Description. Chicago: Society of American Archivists.

Hensen, S. L. (1989). Archives, Personal Papers, and Manuscripts: A Cataloguing Manual for Archival Repositories, Historical Societies, and Manuscript Libraries. (2nd Ed.) Chicago: Society of American Archivists.

Hix D., & Hartson H. (1993). Developing User Interfaces: Ensuring Usability Through Product & Process. New York: John Wiley & Sons.

Holt, C., Kiffer, M. & Peterson, K. (2000). "Transcription and Cataloguing of the Robinson Reports." IQP Worcester Polytechnic Institute.

Johnson, M. (1999). XML for the absolute beginner. In Java World [Online]. Available: http://www.javaworld.com/javaworld/jw-04-1999/jw-04-xml.html [2001 January 30].

Kinder, R. (Ed.). (1994). Librarians on the Internet. New York: The Haworth Press.

Lancaster, F. W. (1978). Toward Paperless Information Systems. New York: Academic Press.

Lancaster, F. W. (1979). Information Retrieval Systems: Characteristics, Testing and Evaluation. New York: John Wiley & Sons.

Lancaster, F. W. (Ed.). (1993). Libraries and the Future – Essays on the Library in the Twenty-first Century. New York: The Haworth Press.

Marsden, C. (2001). (face-to-face interviews, March & April 2001).

Meadow, C. T. (1970). Man-Machine Communication. New York: John Wiley & Sons.

Miller, F. M. (1990). Arranging and Describing Archives and Manuscripts. Chicago: Society of American Archivists.

National Art Library [Online]. Available: http://www.nal.vam.ac.uk/ [2001, February 24].

Oracle Internet. (2000). Using XML in Oracle Database Applications. [Online]. Available: http://technet.oracle.com/tech/xml/info/htdocs/otnwp/about_xml.htm [2001 January 30].

Sall, K. (1998). XML and Java: The Perfect Pair. [Online] Available: http://wdvl.com/Authoring/Languages/XML/Java/index.html [2001 February 11].

Shiffman, H. (1998). Making Sense of Java. [Online]. Available: http://www.disordered.org/Java-QA.html [2001 February 11].

Ullman, J. D., & Widom, J. (1997). A First Course in Database Systems. New Jersey: Prentice Hall.

Valauskas, E., John, N. (Eds.). (1995). The Internet Initiative. Chicago: American Library Association.

Victoria & Albert Museum [Online]. Available: http://www.vam.ac.uk/ [2001, February 24].

Wells, A., Calcari, S., & Kaplow, T. (Eds.). (1999). The Amazing Internet Challenge. Chicago: American Library Association.

Appendix A: NAL Conducted Survey

A.1 Survey Form

NATIONAL ART LIBRARY READER SURVEY

The National Art Library is committed to providing access to all its collections. In order to do so it requires up-to-date information on its users and the Library services they use. Please take the time to answer the following questions to allow the Library to better plan for the future.

Please tick the appropriate boxes.

PERSONAL DETAILS

1. Personal

Male ☐ Female ☐

Age Under 20 ☐ 21-40 ☐ 41-60 ☐ Over 60 ☐

2. Occupation (please tick one box only):

Academic	☐		
Architect	☐		
Artist	☐	Picture researcher	☐
Bookseller	☐	Student (postgraduate)	☐
Curator	☐	Student (undergraduate)	☐
Dealer	☐	Teacher	☐
Designer	☐	Writer	☐
Editor	☐	Other employed	☐
Government	☐	Retired	☐
Journalist	☐	Unemployed	☐
Librarian/Information Scientist	☐	Other (please specify)	☐
Performing Arts	☐		
Publisher	☐	_____	

3. Type of organization employed by (please tick one box only)

Academic institution	☐	Media	☐
Arts organization	☐	Performing arts/entertainment	☐
Auction house	☐	Museum	☐
Commercial	☐	Self-employed	☐
Government department	☐	Other (please specify)	☐
Information industry/publishing	☐		
Legal	☐	_____	
Library	☐		

4. Name of organization (optional)

_____

5. Normal residence (please tick one box only)

London ☐
Other UK ☐
Other EU ☐
North America (USA/Canada) ☐
Elsewhere in world (please specify) ☐

_____

USE OF LIBRARY

6. Purpose of library research (please tick one box only)

Personal ☐
Academic ☐
Work related ☐

7. Frequency of visits to NAL

First visit ☐

Approx. number of visits in last month

_____

Approx. number of visits in last 12 months

_____

Expected number of visits in next 12 months

_____

8. Months in which you visit the library (please tick all that apply)

January	☐	May	☐	September	☐
February	☐	June	☐	October	☐
March	☐	July	☐	November	☐
April	☐	August	☐	December	☐

9. Days of the week you visit the library (please tick all that apply)

Tuesday	☐	Thursday	☐	Saturday	☐
Wednesday	☐	Friday	☐		

10. Times of day you typically visit the library (please tick all that apply)

10.00-12.00	☐	12.00-14.00	☐	14.00-17.00	☐

11. Length of time typically spent (please tick one box only)

Under 1 hour	☐	Up to 5 hours	☐
Up to 3 hours	☐	More than 5 hours	☐

12. Have you ever used any of the following (please tick all applicable)

V & A Archives, Blythe House	☐
Archive of Art & Design, Blythe House	☐
Beatrix Potter Collections, Blythe House	☐
Renier Collection, National Museum of Childhood at Bethnal Green	☐
V & A Departmental Library (please specify)_____	☐

13. Use of other services	**Aware of**	**Ever used**
Telephone enquiry	☐	☐
Telephone requisitions	☐	☐
NAL WWW pages	☐	☐
Postal enquiries	☐	☐
E-mail enquiries	☐	☐
Book reserve at counter	☐	☐
Photocopying	☐	☐
Photography stand	☐	☐

14. Which other libraries do you use (please tick all that apply)

British Library	☐
Tate Gallery Library	☐
National Portrait Gallery Library	☐
Westminster Public Library	☐
Other public libraries (please specify)_____	☐
University, college, research (please specify)_____	☐
Other (please specify)_____	☐

YOUR VISIT TODAY

15. Subject research (please tick all applicable)

Architecture	☐	Librarianship/ Information science	☐
Arts/Crafts/Design	☐	Literature	☐
Biography	☐	Manuscripts	☐
Education	☐	Museum studies	☐
Genealogy	☐	Music	☐
Heraldry	☐	Performing arts	☐
History	☐	Other (please specify)	☐
History of the book	☐		
Bookbinding	☐		
Book art	☐	_____	

16. Main intention (please tick one box only)

To consult material of which you are aware ☐
To search for material unknown to you ☐
Both the above ☐
Other (please specify) _____ ☐

17. Did you request assistance from staff YES ☐ NO ☐

(If YES please tick as applicable)

Subject enquiry ☐
Procedural enquiry ☐
Reader Registration enquiry ☐
Photocopy order ☐
Other (please specify) _____ ☐

18. Material consulted

Library catalogues (microfiche)	☐	Books	☐
Library catalogues (computer)	☐	Periodicals	☐
Library catalogues (printed)	☐	Catalogues (sale/exhibition)	☐
Subject indexes	☐	Microform publications	☐
Databases on CD-ROM	☐	Special Collections Books	☐
Open reference	☐	Special Collections Manuscripts	☐

19. Did you request books for use YES ☐ NO ☐

If YES please tick as applicable

In advance by	telephone	☐
	fax	☐
	letter	☐
	e-mail	☐
At NAL by requisition form	today	☐
	before today	☐

If YES were all items requested available YES ☐ NO ☐

If not, what was the reason given
Out to a member of staff ☐
Not in place on the shelf ☐
Other (please specify) ☐

20. Was your visit successful YES ☐ PARTIALLY ☐ NO ☐

If NO or PARTIALLY please give the reason

_____

21. Will you revisit the Library YES ☐ NO ☐ DON'T KNOW ☐

READER SATISFACTION

22. Rate the Library's services (on a scale of 1 (poor) to 4 (good))

	1	2	3	4
Availability of books and material required	☐	☐	☐	☐
Waiting time for delivery	☐	☐	☐	☐
Time spent queuing	☐	☐	☐	☐
Helpfulness of staff	☐	☐	☐	☐
Staff knowledge	☐	☐	☐	☐
Usefulness of NAL information leaflets	☐	☐	☐	☐
Availability of seats	☐	☐	☐	☐
Working environment	☐	☐	☐	☐
Ease of use of catalogues & indexes	☐	☐	☐	☐
Speed of photocopying service	☐	☐	☐	☐
Opening hours	☐	☐	☐	☐
Cloakroom deposit facilities	☐	☐	☐	☐

23. Agree/Disagree with the following statements on scale of 1 (strongly disagree)-4 (strongly agree)

	1	2	3	4
In my subject the NAL collection is unparalleled	☐	☐	☐	☐
Access to the NAL is essential to my research	☐	☐	☐	☐
The material I use in the NAL is unavailable elsewhere	☐	☐	☐	☐
I use the NAL because it is conveniently located	☐	☐	☐	☐
Other libraries are just as important in my research	☐	☐	☐	☐
I use the NAL because I am most familiar with it	☐	☐	☐	☐
Use of the NAL lends authority to my research	☐	☐	☐	☐
I use the NAL for its quick & reliable service	☐	☐	☐	☐

PLEASE INDICATE WHAT NEW OR EXPANDED SERVICES YOU WOULD LIKE THE NAL TO OFFER:

FURTHER COMMENTS

Date survey form completed

___/__/ 1999

Public Services, National Art Library, Victoria and Albert Museum, Cromwell Road, South Kensington, London SW7 2RL

A.2 Survey Findings

PERSONAL DETAILS

1. Personal

Sex

Male	40
Female	101

Age

Under 20	13
21-40	85
41-60	31
Over 60	15

2. Occupation

Academic	28
Architect	0
Artist	2
Bookseller	3
Curator	6
Dealer	5
Designer	5
Editor	2
Government	1
Journalist	4
Librarian/Information Scientist	2
Performing Arts	0
Publisher	1
Picture researcher	6
Student (postgraduate)	33
Student (undergraduate)	35
Teacher	5
Writer	14
Other employed	5
Retired	10
Unemployed	1
Other (please specify)	8

3. Type of organization employed by

Academic institution	39
Arts organization	5
Auction house	4
Commercial	4
Government department	1
Information industry/publishing	3
Legal	0
Library	0
Media	9
Performing arts/entertainment	2
Museum	8

Self-employed	37
Other (please specify)	15

4. Normal residence

London	102
Other UK	19
Other EU	10
North America (USA/Canada)	6
Elsewhere in world	7

USE OF LIBRARY

5. Purpose of library research

Personal	23
Academic	86
Work related	44

6. Frequency of visits to NAL

First visit	24

Approx. number of visits in last month

0	7
1	12
2	19
3	14
4	10
5	7
5 - 10	12
10+	12

Approx. number of visits in last 12 months

1-5	24
5 - 10	16
10 - 20	15
20+	25

Expected number of visits in next 12 months

1-5	7
5-10	15
10 - 20	18
20+	36

7. Months in which you visit the library

January	82
February	85
March	88
April	88

May	87
June	79
July	68
August	56
September	75
October	91
November	101
December	106

8. Days of the week you visit the library

Tuesday	88
Wednesday	92
Thursday	94
Friday	89
Saturday	75

9. Times of day you typically visit the library

10.00-12.00	85
12.00-14.00	96
14.00-17.0	95

10. Length of time typically spent

Under 1 hour	1
Up to 3 hours	56
Up to 5 hours	54
More than 5 hours	26

11. Have you ever used any of the following

V & A Archives	15
Archive of Art & Design	16
Beatrix Potter Collections	2
Renier Collection, BGM	3
V & A Departmental Library	20

12. Use of other services

Telephone enquiry	45
Telephone requisitions	22
NAL WWW pages	25
Postal enquiries	15
E-mail enquiries	21
Book reserve at counter	56
Photocopying	67
Photography stand	44

13. Which other libraries do you use

British Library	67
Tate Gallery Library	20
National Portrait Gallery Library	48
Westminster Public Library	48
Other public libraries	67
University, college, research	96
Other	42

YOUR VISIT TODAY

14. Subject research Architecture	24
Arts/Crafts/Design	105
Biography	19
Bookbinding	2
Book art	12
Education	1
Genealogy	2
Heraldry	1
History	32
History of the book	7
Bookbinding	2
Book art	11
Librarianship	6
Information science	11
Literature	1
Manuscripts	5
Museum studies	35
Music	1
Performing arts	5
Other (please specify)	35

15. Main intention

To consult material of which you are aware	32
To search for material unknown to you	15
Both the above	89
Other	4

16. Did you request assistance from staff YES 93 NO 37

Subject enquiry	35
Procedural enquiry	62
Reader Registration enquiry	41
Photocopy order	30
Other	4

17. Material consulted

Library catalogues (microfiche)	48
Library catalogues (computer)	110
Library catalogues (printed)	31
Subject indexes	24
Databases on CD-ROM	14
Open reference	25
Books	89
Periodicals	59
Catalogues (sale/exhibition)	45
Microform publications	3
Special Collections Books	31
Special Collections Manuscripts	6

18. Did you request books for use YES 119 NO 16

In advance by:	
Telephone	7
fax	1
letter	0
e-mail	0
At NAL by requisition form	
today	105
before today	21

If YES were all items requested available YES 77 NO 31

If not, what was the reason given

Out to a member of staff	5
Not in place on the shelf	14
Other	16

19. Was your visit successful YES 97 PARTIALLY 28 NO 1

20. Will you revisit the Library YES 126 NO 0 DON'T KNOW 2

READER SATISFACTION

21. Rate the Library's services (on a scale of 1 (poor) to 4 (good))

	1	**2**	**3**	**4**
Availability of books and material required	13	8	32	71
Waiting time for delivery	8	24	39	51
Time spent queuing	8	17	38	57
Helpfulness of staff	10	8	18	78
Staff knowledge	8	7	27	65
Usefulness of NAL information leaflets	4	10	32	32
Availability of seats	9	8	24	83
Working environment	9	9	24	83
Ease of use of catalogues & indexes	11	11	41	52
Speed of photocopying service	11	17	21	17
Opening hours	15	41	38	27
Cloakroom deposit facilities	16	16	36	50

22. Agree/Disagree with the following statements on scale of 1 (strongly disagree)-4 (strongly agree)

	1	**2**	**3**	**4**
In my subject the NAL collection is unparalleled	1	28	50	31
Access to the NAL is essential to my research	5	8	20	87
The material I use in the NAL is unavailable elsewhere	6	9	44	55
I use the NAL because it is conveniently located	20	28	35	36

Appendix B: IQP Team Conducted Survey

B.1 Survey Form

NAL Online Archival Texts Project

In an ongoing effort to provide the best possible access to all its collections, the National Art Library is currently designing a system to place the full text of selected archival documents from the NAL and the V&A Archive on the internet, focussing initially on J.C. Robinson's reports 1863-1868. In order to provide the most complete and useful functionality, we would like to survey those who have worked extensively with these types of documents. Please answer these questions as completely as possible. Thank you for your time.

Name: _____

Department: _____

1. Which of the following archival documents have you used or are you using? Please check all that apply.

   ____ J.C. Robinson's reports 1863-1868
   ____ Art Referees' reports 1868 – 1886
   ____ Abstracts of V&A correspondences 1864 – 1914
   ____ Board minutes of Science & Art Department 1852 - 1892

2. Are there other categories of Museum archives or documents that you can suggest which might be suitable for full-text electronic reproduction?

   _____
   _____
   _____
   _____
   _____

3. Would you find it useful if the text of these documents were available online? Please explain why, or why not, and indicate which texts would be most useful.

   _____
   _____
   _____
   _____

Please continue on other side.

4. How useful would it be if you could have the computer search the documents for the following types of information? (Rate the following on a scale of 1 to 5; 1 = not useful, 5 = extremely useful)

Name of a person	____
Name of a place	____
Name of an object	____
Name of a material used	____
Name of an organisation	____
Name of a museum collection	____
Cost or expense of a purchase	____
Specific date	____
Range of dates	____
Museum object number	____
Other (please specify):	
_____	____
_____	____
_____	____

5. Do you have any comments or suggestions about the creation of this online system? We are most interested in other features that you would like to see, such as the inclusion of links to pictures of art objects or different methods of display formatting.

_____

_____

_____

_____

_____

_____

_____

6. If you have 10 to 15 minutes available for either a phone or a face-to-face conversation, we would greatly appreciate the chance to contact you. If you are interested, please write your contact information below and we will schedule a time when it is most convenient for you.

Phone: _____

Email: _____

Please return this survey to Mr. Christopher Marsden by April 6th, 2001.

B.2 Survey Findings

Between the dates of April 1st and April 20th 2001, five surveys were returned.
The following is a summary of the responses that we received.

1. Which of the following archival documents have you used or are you using? Please check all that apply.

Documents	Number of people who used them
J.C. Robinson's reports 1863 – 1868	3
Art Referees' reports 1868 – 1886	5
Abstracts of V&A correspondences 1864 – 1914	2
Board minutes of Science & Art Department 1852 – 1892	4

2. Are there other categories of Museum archives or documentation that you can suggest which might by suitable for full-text electronic reproduction?

"Ideally, an electronic index of all entries in the entries in the Card Catalogue in the Furniture Information Section. This would be very labour intensive."

"Nominal Files relating to major acquisitions or collectors. Art Museum inventories – especially 1851 volume."

"Annual reports of the Museum. Inventories of Museum objects (although this information should mainly appear online already in the Collections Information System)."

3. Would you find it useful if the text of these documents were available online? Please explain why, or why not, and indicate which texts would be most useful.

"Yes – easier access and all that it involves – no need to make appointments in the NAL and block off time to see papers, when one has other heavy commitments."

"Yes, but handwriting gives a lot of information so there should be some indication of the author of any anonymous notes, if recognisable from writing and not otherwise identified."

"Yes, hopefully easier access."

"Online access would clearly be advantageous especially for enabling remote access for members of the public. It would also allow speedier access to core documentation for members of staff, and save wear and tear on original documents. The value of the online surrogate would partly depend on the accuracy of transcription and on ensuring data, particularly in complex-structured documents, did not become distorted by being displayed out of context. The Board Minutes, which are printed and simply structured, would be the easiest to transcribe and display reliably; the earliest volume

is also particularly useful in documenting the 1850s for which there are few other records."

"While I think of it – re Robinson papers and Art Referees reports: I have found these extremely useful in the past, and an index will be invaluable, if done thoroughly."

4. How useful would it be if you could have the computer search the documents for the following types of information? (1 = not useful, 5 = extremely useful)

Search Category	**People ranked them**				
	1	**2**	**3**	**4**	**5**
Name of a person	-	-	-	1	3
Name of a place	-	-	1	2	1
Name of an object	-	-	-	3	1
Name of a material used	1	-	1	2	-
Name of an organisation	-	-	1	1	2
Name of a museum collection	-	1	-	2	1
Cost or expense of a purchase	-	2	-	1	1
Specific date	-	-	-	2	2
Range of dates	-	-	1	1	2
Museum object number	-	-	-	-	4
Other (please specify):					
Registered Paper number	-	-	-	1	1

5. Do you have any comments or suggestions about the creation of this online system? We are most interested in other features that you would like to see, such as the inclusion of links to pictures of art objects or different methods of display formatting.

"Pictures – v. useful."

"Links to images would be very useful. All these sources are little used in our department [Indian] – it is probably true to say that most people have never used all of them, and one or two may never have used any of them. This is probably due in part to not knowing where everything is, or not knowing the sort of information that can be gleaned from the different parts of the archive. It might be useful in the interim to issue a short note outlining the contents – this would be very useful for new members of curatorial staff. For others who do know the treasures they contain, having online access to the contents would certainly be helpful, if the cost of doing this is not absolutely prohibitive."

"Links to the collections of information system would be most useful; which would in turn be linked to images of the Museum objects. Running headers which indicate to the user at all times the title, reference number and dates of the document being viewed would be helpful. For complex documents such as the Robinson/Art Referees' reports a tabulated display layout would be necessary, but also a facsimile

layout, which presents data as far as possible in context – unless an actual facsimile image of the document is also going to be linked to each entry."

"I would be interested to know more about how the texts are to be digitised e.g. plain text, xml, scanning etc and how retrieval mechanisms will work. Also, there are connections that could be made with other types of material held around the Museum and I would be interested to know what the current thinking is on making these relationships."

"I would like to be able to search on material (e.g. bronze, terracotta) and on subject matter, as well as names of people, cities and so on."

Appendix C: IQP Team Analysis of Tag Sets

This analysis of the contextual tags of the TEI and EAD tag sets includes a definition of each tag and the tag's attributes. These are the tags and attributes that will be used to implement the search system.

C.1 TEI Tag Set

<tag> - definition
"attribute" - definition

<abbr> - An abbreviation
"expan" - The expanded form of the abbreviation
"type" - The type of abbreviation, such as acronym, or contraction
"cert" - The certainty of the transcriber that the expansion is correct

<biblScope type = superceded> - The Registered Paper number of the physical document
"type" - This attribute is permanently set to "superceded"

<collectionName> - The name of a Museum collection.
"reg" - The regularised or standardised form of the name

<cost> - The cost of a transaction, such as the purchase of an artefact.

<date> - A specific, singular date
"value" - The standardised version of the date, in YYYY-MM-DD format.
"certainty" - The certainty the transcriber has that the date given is correct, on a 0 to 100 scale, with 100 being absolutely sure the date is accurate

<dateRange> - A span of time, such as a specific mention of two dates or a general time range, such as "15th Century."
"from" - The standardised version of the beginning date of the range
"to" - The standardised version of the ending date of the range
"exact" - The accuracy of the given range, used if the transcriber estimated the "from" and "to" values because the range given wasn't specific

<eventName> - The name of a notable event
"reg" - The regularised or standardised form of the name

<materialName> - The name of a material that an art object is made of
"reg" - The regularised or standardised form of the name

<name type = site> - The name of a specific site, such as a train station or building.
<orgName> is used to represent names such as the V&A Museum, which is known more as an organisation than as the name of the building
"reg" - The regularised or standardised form of the name
"type" - This attribute is permanently set to "site"

<objectIdentifier> - The Museum Object number, a unique number assigned to each object at the V&A
"reg" - The standardised (current) form of the Museum Object number, it has changed format over the years

<objectName> - The name of an art object
"reg" - The regularised or standardised form of the name

<orgName> - The name of an organisation of any sort
"reg" - The regularised or standardised form of the name

<persName> - The name of a person
"reg" - The regularised or standardised form of the name
"type" - The relationship of the person to the document, such as author, subject, or photographer

<placeName> - The name of a place, such as countries and cities, or natural geographical features such as mountains or rivers
"reg" - The regularised or standardised form of the name

<orig> - Antiquated text, words whose spellings have changed over time
"reg" - The modern spelling of the antiquated word

<sic> - A spelling mistake on the physical document
"corr" - The correct spelling of the word
"cert" - The transcriber's certainty that the corrected spelling is accurate, on a 0 to 100 scale

<supplied> - Text that has been supplied by the transcriber if there is damage or gaps in the text
"reason" - The reason why the text was supplied, i.e. what had happened to the physical document

<title> - The title of the document

C.2 EAD Tag Set

<tag> - definition
"attribute" - definition

<abbr> - An abbreviation within the text
"expan" - The expanded form of the abbreviation

<corpname> - The name of a corporation, company , or other professional group
"normal" - The standardised form of the name
"role" - The relationship between the name and the document

<date> - A date or a range of dates within the document
"normal" - The standardised form of the date, written as YYYYMMDD for a single date or YYYYMMDD – YYYYMMDD for a range
"type" - A more specific reason for the date, such as publication or birthday
"certainty" – An indication of whether the date is meant to accurate or an approximation such as "circa"

<expan> - The expanded form of a common abbreviation
"abbr" - The abbreviated form of the expansion

<famname> - The name of a family or a part of a family group
"normal" - The standardised form of the name
"role" - The relationship between the name and the document

<geogname> - The name of a place or location
"normal" - The standardised form of the name
"role" - The relationship between the name and the document

<name> - A general tag for names, used when the more specific tags is inapplicable
"normal" - The standardised form of the name
"role" - The relationship between the name and the document

<persname> - The name of a specific person
"normal" - The standardised form of the name
"role" - The relationship between the name and the document

<title> - The title of a work mentioned within the text, such as an art object or book
"normal" - The standardised form of the title

<unitdate> - The date on which the physical document was created
"normal" - The standardised form of the date, written as YYYYMMDD

<unitid type = previous> - The Registered Paper number of the physical document
"type" - This attribute is set to "previous" to separate the RP number from the other types of document identification

<unittitle> - The title of the document

Appendix D: Database Design

D.1 Main Table Tags

The following tags each have their own table in the database, as they are each searchable options:

TEI:

<biblScope type = superceded>
<collectionName>
<cost>
<date>
<dateRange>
<eventName>
<materialName>
<name type = site>
<objectIdentifier>
<objectName>
<orgName>
<persName>
<placeName>

EAD:

<corpname>
<date>
<famname>
<geogname>
<name>
<persname>
<title>
<unitdate>
<unitid type = previous>

D.2 Subtable Tags

The following tags will only appear as subtables, as they are only included in searches when they are nested in one of the main searchable tags.

TEI:

<abbr>
<orig>
<sic>
<supplied>

EAD:

<abbr>
<expan>

D.3 Database Table Objects

Each type of tag has a database object assigned to it. An object is a standard table layout for the particular tag. As the tables are designed to grow dynamically, the database objects are very important. When a tag is nested, the database will retrieve the appropriate object and use it to format the subtable, which places the information of nested tags within the main tag. This dynamic expansion saves the programmer time, as he or she does not have to create each new table manually. The database will create and expand tables automatically as required.

D.3.1 TEI Database Objects:

abbreviation object – applies to the <abbr> tag, exists only as a subtable

abbr				
text	expan	type	cert	subtables

antiquated text object - applies to the <orig> tag, exists only as a subtable

orig		
text	reg	subtables

date object - applies to the <date> tag

date			
text	value	subtables	link to document

date range object - applies to the <dateRange> tag

dateRange				
text	from	to	subtables	link to document

cost object - applies to the <cost> tag

cost		
text	subtables	link to document

Museum Object number object - applies to the <objectIdentifier> tag, there are no possible subtables

objectIdentifier		
text	reg	link to document

name object - applies to all of the TEI name tags

*Name				
text	reg	type	subtables	link to document

Registered Paper number object - applies to the <biblScope type = superceded> tag, there are no possible subtables

biblScope		
text	type = superceded	link to document

spelling mistake object - applies to the <sic> tag, exists only as a subtable

sic			
text	corr	cert	subtables

supplied text object - applies to the <supplied> tag, exists only as a subtable

supplied		
text	reason	subtables

D.3.2 EAD Database Objects:

abbreviation object - applies to the <abbr> tag, exists only as a subtable, there are no possible subtables

abbr	
text	expan

date object - applies to the <date> tag, there are no possible subtables

date				
text	normal	type	certainty	link to document

expansion object - applies to the <expan> tag, exists only as a subtable, there are no possible subtables

expan	
text	abbr

name object - applies to all EAD name tags

*name				
text	normal	role	subtables	link to document

Registered Paper number object - applies to the <unitid type = previous> tag, there are no possible subtables

unitid		
text	type = previous	link to document

title object - applies to the <title> tag

title			
text	normal	subtables	link to document

unit date object - applies to the <unitdate> tag, there are no possible subtables

unitdate		
text	normal	link to document

D.4 Example of a Database Table

This is an example of how a database table would function with a piece of tagged text.

Document text: Mr. D. Wells

TEI tagged text:

<persName reg = Wells, Digby type = recipient> <abbr expan = Mister type = common abbreviation cert = 100> Mr. </abbr> D. Wells </persName>

Database table:

persName								
text	reg	type	subtables					link to document
			abbr					
			text	expan	type	cert	subtables	
D. Wells	Wells, Digby	recipient	Mr.	Mister	common abbreviation	100		link.xml

Because an abbreviation tag was nested within the name tag, the abbreviation object was entered into the subtable area of the name table. The persName table contains all of the information represented by the tagged text.

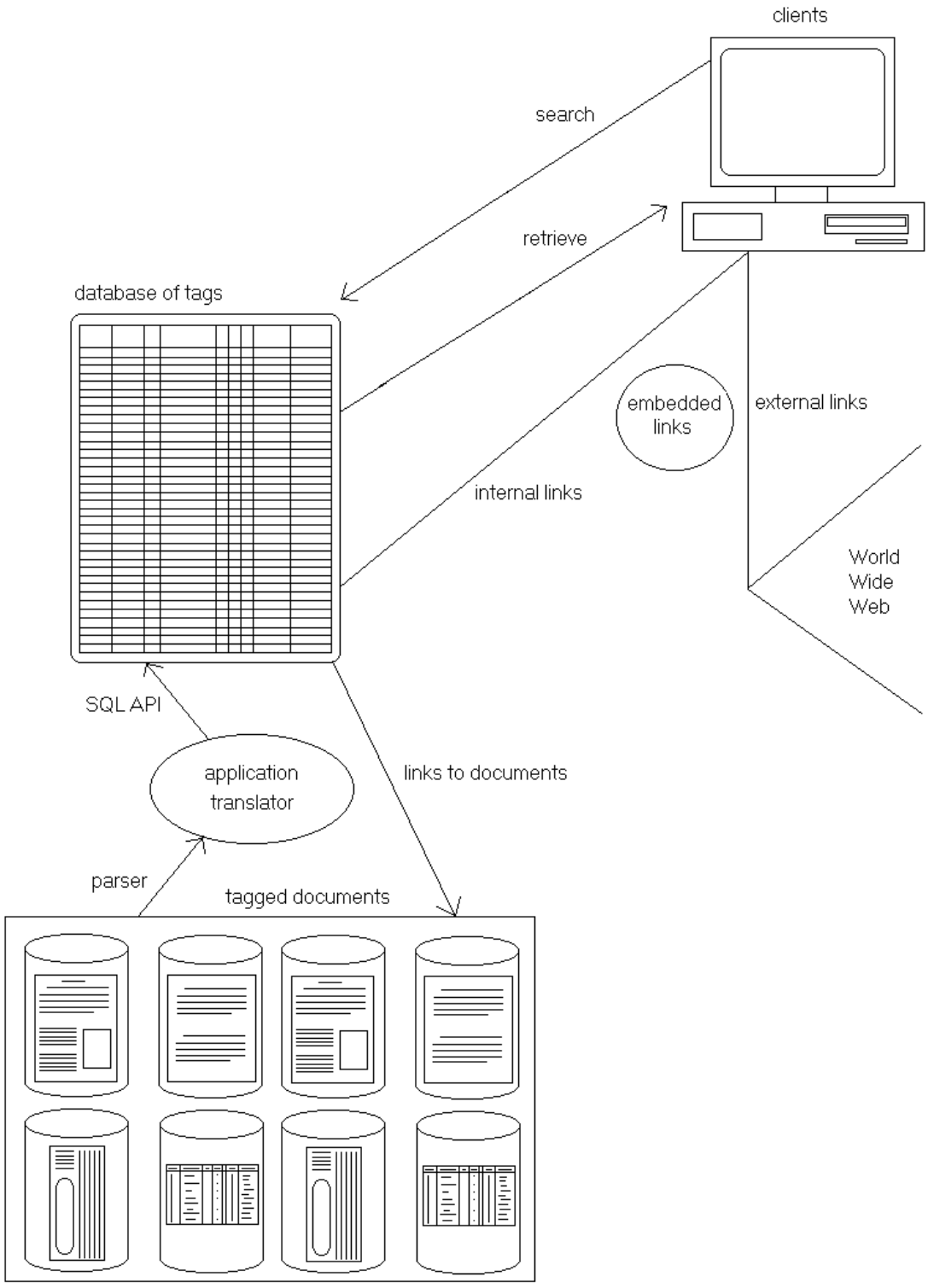
D.5 Database Search Engine Concerns

Several algorithms will need to be written by the database programmer to expand the capabilities of the search engine. First, a closest probable match algorithm is necessary to find words that are similar to the search parameters but not exact. Similarly, an algorithm will need to be created to translate dates that are entered into the search field into the standardised forms used for searching the date tags. The search engine will also need to understand date and cost ranges and be able to find values that fall between the range limits.

Appendix E: Interface Prototype

Please refer to the Visual Basic files prototype.vbp and prototype.vbg, or the executable file prototype.exe.

Appendix F: High Level System Overview



Appendix G: Glossary

AACR 2 – Anglo-American Archiving Rules, a set of standard adopted by manuscript curators, archivists, and librarians to abide by when archiving materials.

archiving – The process of organizing and storing items.

applet – A small computer program transferred to a user's computer via the internet.

browser – A type of computer software (such as Netscape, Mosaic and Internet Explorer) that allows a user to view webpages.

cataloguing – The process of organizing materials according to a specific scheme such as chronological, chronological by subject or original.

CD-ROM – Compact disk read-only media, a form of optical media that is portable, durable and can store large amounts of data.

CGI – Common Gateway Interface, a type of programming language used in online systems that is maintained on the server, rather than downloaded to the user's computer.

database – A collection of information managed by a database management system (DBMS); can be made accessible via the internet, in which case it is referred to as an online database.

data-definition language – A class of computer languages used to create electronic databases.

EAD – Encoded Archival Description, a tag set that used for electronically transcribed abstracts. See also *TEI*.

E/R model – Entity-relationship models, one of the two major standards for data-definition languages (along with ODL models). Composed of entities with two properties: attributes and relationships.

embedded links – Links within the text of online documents indicated with keywords.

HCI – Human Computer Interaction, what happens when a human user and a computer system get together to perform tasks, includes such diverse issues as user interface software and hardware, user and system modelling and cognitive and behavioural science.

HTML – Hypertext markup language, a simple markup language designed to give instructions on how to format a page; the language used to create most webpages.

ILL – Inter-library loan, the loaning of books from one library to another.

integrity – In terms of computers, this refers to the security of a system from unwanted access.

IPL – Internet Public Library, one of the largest and most popular online libraries in the world (located at http://www.ipl.org/).

IQP – Interactive Qualifying Project, one of three projects in the "WPI Plan" required for graduation.

Java – A modern programming language largely used as a means to run small, online programs.

JavaScript – A scripting language completely unrelated to Java that does not have to be compiled on every computer it runs on, often used for small online programs.

keyword – A word, or group of words, that the transcriber has marked with tags to indicate importance.

markup language – A computer language used to format text, predominantly used to create webpages.

metadata – Information stored in an electronic database that indicates how the data is stored.

microfilm – A form of publication and reproduction that used small cameras and film to make large collections available to libraries.

Museum Object Identifier – A unique number assigned to every Museum object for identification and reference purposes.

NAL – National Art Library, a division of the V&A Museum where we will be working.

ODL – Object definition language; along with E/R models, one of the two major standards for data-definition languages. Composed of objects with three properties: attributes, relationships and methods.

online database – See *database*.

optical media – Computer hardware (such as CD-ROMS and hard drives) that are used to store large amounts of information.

parser – Computer code designed to break up information into its component pieces; for XML, involves separating the document by its tags and thus revealing the structure and attributes of the tags and the document as a whole

Perl – A programming language used for text handling and manipulation that has evolved into a complex programming language capable of almost any task.

photostat machine – A precursor to modern photocopy machines. See *Xerox 914*.

query processor – The part of a DBMS that handles input to the database.

Registered Paper Number – A method the Museum uses to label archived material by stamping each page with a unique number. This number is an indication of the chronological order in which the papers were filed into the archives, although it does not always follow the chronological order in which they were written

Robinson Reports – A series of letters written by Sir John Charles Robinson to the Victoria & Albert Museum in the mid to late 19th century largely concerning art acquisitions.

schema modifications – A type of input to a DBMS that changes the layout of data storage.

SGML – Standard Generalized Markup Language, the first markup language designed to format a document or webpage

SQL – Also know as sequel, one of the most common database manipulation languages; all SQL queries follow the same "select-from-where" structure.

storage manager – The part of the DBMS that is responsible for obtaining information from the data storage and modifying the information when requested by the levels of the system above it.

style sheets – An omnipresent extension of any markup language that defines how to display the tag elements.

tagging – the process of adding tags to electronic documents. Tags are used to format text (such as separating paragraphs and italicising, as well as indicating keywords.

TEI – Text Encoding Initiative, an electronic organisation method used to maintain the documents in the system that the NAL set up for the Robinson Reports.

transaction manager – The part of the DBMS that ensures system integrity.

transcription – The process of converting hand written documents to an electronic format.

V&A – Victoria & Albert Museum, the sponsoring agency for this project.

Xerox 914 - The first widely used copy machine that allowed users to quickly, easily and cheaply copy library materials for off-site use.

XML – A markup language used to define new markup languages, allowing one create a language crafted for a specific application or domain; gives a programmer more freedom with the language and complements existing HTML while allowing more options and the ability to step outside HTML's boundaries.

XSL – The style sheet the NAL chose to use for its documents; a multifaceted style sheet with numerous functions for controlling document displays, XSL sheets can translate XML into HTML, or into a different XML dialect.