LRN: ~~00D070I~~
00E0341
CXD-LON2-45

TRANSCRIPTION AND CATALOGUING OF THE ROBINSON REPORTS

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

Chris Holt

Micah Kiffer

Keith Peterson

Date:

Approved:

Professor Richard Vaz, Co-Advisor

Professor Chrysanthe Demetry, Co-Advisor

Abstract

The National Art Library in London possesses a collection known as the Robinson Reports that is inaccessible to scholars and difficult to read and search. To address these problems, Professor Lee Fontanella of WPI will be transcribing the reports. The goal of this project was to develop aids for his work and to facilitate the development of an electronic search system for the transcriptions. We developed an editing environment to store the transcriptions in a non-proprietary format and a means of tagging information in the reports for subsequent searching. We also prepared a user manual for the transcription process and documented our recommendations for the development of the electronic search system.

Executive Summary

Sir John Charles Robinson was the primary art collector for the Victoria and Albert Museum (V&A) in London from 1853-1868. During this time he travelled Europe, searching out and buying works of art for the museum's collection. Robinson was in constant contact with the museum, reporting on what he had found and what he needed to obtain it. His letters and associated documents, known as the Robinson Reports, are housed within the National Art Library (NAL) at the V&A and provide a great deal of information on the provenance of many of the museum's holdings; they also offer valuable insight into Robinson's life and the history of the museum. Just one of the valuable and interesting pieces of information his reports offer is a unique first-person account of his unceremonious dismissal.

Although information in the Robinson Reports is of great interest to scholars, it is quite difficult for them to use the reports in their research. This is not an uncommon occurrence with old and valuable documents. Museums, libraries, and collectors – the NAL included – guard documents like the Robinson Reports very carefully. Documents of this nature are very fragile and irreplaceable; because of this, scholars need to get special permission to work with them. Even when these documents are available, they are often very hard to search through for specific information; this is because older documents are generally catalogued very poorly, or not catalogued at all. A lack of cataloguing forces scholars to painstakingly read through each document to find the information for which they are looking. For instance, if an art scholar were looking for pieces that Robinson purchased by Donatello, that scholar would need to read every report in the collection and choose items by hand that discuss Donatello. In addition,

handwriting styles have changed a great deal over the years, making it hard to read older texts. To compound this problem, care was not always taken when writing documents. The Robinson Reports in particular were often written very quickly, causing them to be quite difficult to read.

These difficulties have spurred international interest in making old and valuable written materials available in an electronic format. This would allow documents to be made available on-line, which would provide access to anyone. At the same time, it would make them easy to read and save scholars time in trying to decipher them. Finally, computer software can be implemented that would permit scholars to search collections of documents for specific information.

However, the differences in software and formats used for providing documents in electronic forms pose a problem when institutions and organisations wish to share information. If they employ different software or formats their data may be incompatible. This problem has sparked interest in the development and implementation of standards for specifying electronic formats that can be read by multiple software packages.

The NAL, along with many other organisations, has taken an interest in electronic transcription via non-proprietary standards. They are currently working on a project for their online catalogue using the Encoding for Archival Description (EAD) standard, a non-proprietary markup language. Professor Lee Fontanella, a Robinson scholar and faculty member at WPI, will be transcribing the Robinson Reports in the summer of 2000. These transcriptions will be stored in a non-proprietary format defined by the Extensible Markup Language (XML), following closely the standards set by the Text

Encoding Initiative (TEI) and the Consortium for the Computer Interchange of Museum Information (CIMI). TEI is an organisation dedicated to developing a standard way for making documents available to a broad audience in an electronic format. CIMI has a similar goal, in developing a standard electronic format for museum information. These standards are based on tags within the text that mark important information regarding the structure and content of the document. These tags can be interpreted by software applications, and presented any way in which the system designer sees fit.

So that it is unnecessary for Professor Fontanella and other scholars to become familiar with technical details, it was our goal to facilitate the transcription process by developing technical aids. These aids are meant to help him understand and use the standards to mark up the Robinson Reports while he transcribes them. Additionally, we prepared for the development of an electronic search system that would be able to locate letters containing specified information.

We had to meet several objectives to accomplish these goals. We had to select an XML environment for Fontanella to use, choose and compile a subset of the TEI and CIMI standard tags that was best suited to the Robinson Reports, and develop a template for each of the different types of documents Fontanella will be transcribing. In addition, we decided to develop a detailed user manual for Fontanella and a guide containing our recommendations for completing the system.

To complete these objectives, we followed several distinct methodical steps. We began by researching Robinson and his letters, relevant cataloguing techniques, electronic systems, and the Standard Generalized Markup Language. This research was applied to the various aspects of our work. Building from this background research, we began

working toward achieving our goals. Our first step was interviewing cataloguers at museums and libraries in Worcester, Massachusetts, to develop an idea of how cataloguing systems look and operate.

Once we had an impression of the system we would be working towards, we began working with the Robinson Reports themselves. We conducted an in depth analysis of the content of the reports, identifying key information to be tagged, unique formats and styles for the development of transcription templates, and common phrases that could have shortcuts defined in the templates.

The next step we took was to develop a complete tag set for the transcriptions. We compared the key information we identified in our analysis of the Robinson Reports to the TEI and CIMI standards. Then we combined a subset of the TEI tags and a few crucial tags from the CIMI standard to create our final tag set. This set was comprised of the TEI core tag set, the prose base tag set, five additional tag sets, and six CIMI tags. The core tag set is mandatory for any TEI compliant text and provides general tags that can be used by all types of documents. The prose base tag set implements a part of the core tag set that defines the structure of the actual text. The additional tag sets contain tags used for more specific applications that would not be necessary in all types of documents; this includes information specific to editorial marks and interpretational notes common in transcriptions. The tags imported from the CIMI standard deal with information specific to museums that is unavailable in TEI.

We used this tag set to develop templates based on the unique formats and styles that we identified in our analysis of the reports. Each template defines the structure for one of the document types, so that Fontanella will have a consistent format for the

transcriptions. They also incorporate shortcuts to the common phrases that we found in our analysis. These templates prompt the user for information that appears regularly in the associated document type; however, the content of the reports, along with any minutes and annotations, must be entered manually.

Professor Fontanella has no knowledge of XML, TEI, or CIMI; to provide him with a means of understanding how to work within an XML environment to transcribe the Robinson Reports, we developed a user manual. The user manual contains a detailed description of all the tags, an explanation of how to use the basic features of the XML editor and document templates, and guidelines to follow when marking up the documents. Sample transcriptions are also included in the user manual to serve as a reference for Fontanella. To ensure that it was clear and complete, Professors Demetry and Vaz of WPI used the manual to transcribe several of the reports.

A separate guide about completing the search system is also provided. This guide describes the methods used to mark up the transcriptions, a detailed explanation of the tags, and recommendations on what can be done in order to utilise the full potential of the marked up electronic versions of the Robinson Reports.

After the search system is completed, the efficiency with which scholars will be able to do research into the letters will be significantly improved. This system will provide new ways in which scholars will be able to use these extremely important and interesting letters. Because the transcriptions will be encoded in a standard non-proprietary format, they will be easily interpreted by many software applications, which will make them accessible to a wider range of scholars and organisations. In addition, the work done with the Robinson Reports will advance future research into standard formats

for transcribing and encoding documents. As more collections are made available electronically, the range and efficiency of scholarly research will be greatly increased. Furthermore, the tag set that we developed is extremely broad and may be applicable and transferable to similar art history archives around the world.

Table of Contents

ABSTRACT ..... **I**

EXECUTIVE SUMMARY ..... **II**

TABLE OF FIGURES ..... **XI**

TABLE OF TABLES ..... **XI**

1 INTRODUCTION 1

2 LITERATURE REVIEW 5

2.1 SIR JOHN CHARLES ROBINSON ***5***
2.2 THE ROBINSON REPORTS ***8***
2.3 CATALOGUING ***10***
2.3.1 Library Cataloguing 11
2.3.1.1 Subject Headings *12*
2.3.1.2 Classification Systems *14*
2.3.2 Archiving 15
2.3.3 Information Retrieval and Electronic Cataloguing 16
2.3.3.1 MARC Formats *17*
2.3.3.1.1 USMARC Formats 17
2.3.3.1.2 USMARC AMC and ArtMARC 19
2.3.3.2 Subject Indexing *20*
2.3.3.2.1 Automatic Indexing 20
2.3.3.2.2 Using Subject Indexing for Retrieval 21
2.4 STANDARD GENERALIZED MARKUP LANGUAGE (SGML) ***22***
2.4.1 SGML Document Structures (DTD) 23
2.4.1.1 Elements and Attributes *24*
2.4.1.2 Entities *24*
2.4.2 SGML Document 25
2.4.3 Advantages of SGML 26
2.4.3.1 Descriptive Markup Language *27*
2.4.3.2 Document Type Definition *28*
2.4.3.3 Portability *28*
2.4.4 Extensible Markup Language (XML) 28
2.4.4.1 HTML Limitations *28*
2.4.4.1.1 Non-Extensible 29
2.4.4.1.2 Display-Centric 29
2.4.4.1.3 Reusability 29
2.4.4.1.4 One "View" of the Data 30
2.4.4.1.5 Semantic Structure 30
2.4.4.2 Uses of XML *30*
2.4.4.2.1 Customising the Presentation of Data 31
2.4.4.2.2 Exchanging Business Data Among Applications 31
2.4.5 Text Encoding Initiative (TEI) 31

2.4.5.1 TEI Schema *32*
2.4.5.2 Consortium for the Computer Interchange of Museum Information *33*

3 METHODOLOGY **35**

3.1 DEVELOPMENT OF GOALS AND OBJECTIVES ***35***
3.2 INTERVIEWS OF CATALOGUERS ***38***
3.3 SELECTION OF EDITING SOFTWARE FOR TRANSCRIBING THE REPORTS ***39***
3.4 ANALYSIS OF THE ROBINSON REPORTS ***41***
3.4.1 Unique Formats and Styles 41
3.4.2 Information to be Tagged 41
3.4.3 Common Phrases 43
3.5 THE DEVELOPMENT OF TAGS ***44***
3.5.1 Comparing the Initial Tag List with the TEI and CIMI Standards 44
3.5.2 Finalising the Tag List 46
3.5.3 Defining the Document Type Definition (DTD) 47
3.6 CREATION OF TRANSCRIPTION TEMPLATES ***49***
3.6.1 Determining Document Types 50
3.6.2 Developing the Templates 50
3.7 SAMPLE TRANSCRIPTIONS OF THE ROBINSON REPORTS ***51***
3.8 DEVELOPMENT OF A USER MANUAL FOR TRANSCRIBING THE REPORTS ***53***
3.9 FINAL TESTING ***54***
3.10 ESTIMATION OF TRANSCRIPTION TIME ***55***
3.11 DEVELOPMENT OF GUIDELINES FOR COMPLETING THE SEARCH SYSTEM ***56***

4 RESULTS AND ANALYSIS **58**

4.1 ANALYSIS OF THE ROBINSON REPORTS ***58***
4.1.1 Unique Formats and Styles 58
4.1.2 Information to be Tagged 61
4.1.3 Common Phrases 66
4.2 SELECTION OF AN EDITING ENVIRONMENT ***67***
4.3 FINAL TAG SET ***68***
4.3.1 Tags Necessary for Marking Up the Text 70
4.3.1.1 Structure and Grammar *70*
4.3.1.2 Lists *71*
4.3.1.3 Physical Appearance of Text *71*
4.3.1.4 Internal Editorial Changes *72*
4.3.1.5 External Editorial Changes *72*
4.3.1.6 Text Style and Tone *73*
4.3.1.7 Names, Dates, and Numbers *73*
4.3.2 Tags Used Only in Document Templates 74
4.3.2.1 The TEI Header *75*
4.3.2.1.1 The File Description 76
4.3.2.1.2 The Profile Description 77
4.3.2.2 The Text *77*
4.3.3 Tags Used for Future Markup of Transcriptions 78
4.3.3.1 Components of Names of People *78*
4.3.3.2 Components of Names of Places *79*

4.3.3.3 *Components of Names of Organisations* ... 80
4.3.3.4 *Inserting Images* ... 80
4.3.3.5 *Linking* ... 81
4.4 *Transcription Templates* ... 81
4.5 *Sample Transcriptions* ... 82
4.6 *Estimates of Transcription Time* ... 82
5 PROSPECTS AND RECOMMENDATIONS ... 84
6 BIBLIOGRAPHY ... 87
7 APPENDIX A – THE VICTORIA AND ALBERT MUSEUM ... 89
8 APPENDIX B – EXAMPLES OF THE ROBINSON REPORTS ... 91
8.1 *30 May 1866* ... 91
8.2 *6 September 1866* ... 91
8.3 *23 October 1866* ... 92
8.4 *18 December 1866* ... 92
9 APPENDIX C – INTERVIEW QUESTIONS ... 93
9.1 *Interview of Professor Fontanella* ... 93
9.2 *Interviews at Libraries and Museums* ... 94
10 APPENDIX D – THE TEI EXTENSION FILES ... 95
11 APPENDIX E – UNNECESSARY TAGS IN THE TAG SET ... 111
11.1 *Document Structure* ... 111
11.2 *Bibliographic Citation* ... 112
11.3 *Title Pages* ... 113
11.4 *Glossaries and Indices* ... 114
11.5 *Editorial Changes to the Text* ... 115
11.6 *Fiction, Verse and Drama* ... 116
11.7 *Names, Addresses and Formulae* ... 116
11.8 *Dates and Times* ... 117
11.9 *Tables and Lists* ... 118
11.10 *Segmentation* ... 118
11.11 *Linking* ... 119
11.12 *Interpretation* ... 119
12 APPENDIX F – USER MANUAL FOR THE TRANSCRIPTIONS ... 121
13 APPENDIX G – GUIDELINES FOR COMPLETING THE SYSTEM ... 122

Table of Figures

Figure 1 – Search and Retrieval Model (Vickory 80) 21

Table of Tables

Table 1 – Definitions of the first digit of MARC codes (Furrie 9) 17

Table 2 – Document Types from the Analysis of the Robinson Reports 59

Table 3 – Key Information from the Analysis of the Robinson Reports 62

Table 4 – Common Phrases from the Analysis of the Robinson Reports 66

Table 5 – Common Phrases from the Analysis of the Robinson Reports 67

1 Introduction

Sir John Charles Robinson was an art collector and a diplomat, capable of finding and obtaining great works of art. Because of this skill, he was appointed primary art collector in the early years of the Victoria and Albert Museum (V&A). Between the years 1853 and 1868, Robinson contributed heavily to the amassing of the world's largest collection of the decorative arts. However, his eccentric personality and high ideals about art and the role of museums caused his relationship with the V&A to deteriorate and Robinson was forced out of the museum in 1868.

While he was in their employment, Robinson wrote hundreds of reports to the museum recounting his journeys. To this day, the National Art Library (NAL) in the V&A holds Robinson's reports, along with those of his successors. The Robinson Reports contain important information regarding the provenance of many of the museum's holdings. In addition, they give an incredible amount of insight into his life and journeys, including a unique first-person account of his unceremonious dismissal.

Although they are rich with information, scholars find it hard to use the Robinson Reports in their research. Due to their age and value, the NAL only allows scholars with special permission to use them, and even then they must work with the letters through mylar sleeves. The restricted availability of the reports can hinder scholars' research.

Even after gaining access to the reports, there is no efficient way for scholars to search through them and find information pertinent to their research. If scholars were attempting to find information on a particular artefact or a collection of artefacts, they would need to know when the artefact or collection was reviewed or purchased by the

museum. Even with this knowledge, they may need to search through several binders of reports because some of the letters are not in chronological order.

To further compound the problem, it is very difficult for scholars to read the Robinson Reports. J. C. Robinson had a writing style that was characteristic of his time, which can often make the letters difficult to read. In addition, many of the letters were written very quickly and with little care. This can make the process of deciphering them very painstaking and sometimes nearly impossible for the untrained eye.

These problems are not exclusive to the Robinson Reports; many museums, libraries, and collectors restrict access to documents of this nature. Documents like the Robinson Reports are irreplaceable and often very fragile; therefore, they cannot be made publicly available. In addition, older documents are not usually included in cataloguing systems, which makes it difficult for scholars to locate them.

As a result of the hardships involved in using these sources, there has been a growing interest in using current technology to make these documents available electronically. The latest networking capabilities permit electronic documents to be made available through the web or on museums' and libraries' local networks. This allows scholars to locate and readily access important sources of information for their research.

However, differences in software and formats used for providing documents in electronic form pose a problem when institutions and organisations wish to share information. These differences may cause their data to be incompatible. This problem has sparked interest in the development and implementation of standards for specifying electronic formats that can be read by multiple software packages. The NAL is among the institutions that have developed an interest in the use of these standards.

Recently, Professor Lee Fontanella, a Robinson scholar and WPI faculty member, asked and received permission from the NAL to transcribe the Robinson Reports to ease his research. These transcriptions will be in a format defined by the Extensible Markup Language (XML), following closely the standards set by the Text Encoding Initiative (TEI) and the Commission for the Computer Interchange of Museum Information (CIMI). XML is a non-proprietary format; therefore the transcriptions will not be locked into a single software application, but rather will be able to be shared amongst all applications that can interpret XML.

After Fontanella's transcriptions are completed, an electronic search system will be developed that can interpret the XML and display the letters to anyone who is interested in the Robinson Reports. Making the system easily accessible, possibly via the web, will allow scholars from around the world to use the reports without having to travel to the NAL. By allowing scholars to search according to specified criteria, the search system will facilitate research into the reports or the artefacts to which they refer. Some of these criteria may include the chronology and classification of the artefact, its history and provenance, or the geographical areas where it originated and was purchased. This system will assist scholars in their research and provide a foundation into which future research into the letters can be added and updated.

The goal of this project was to develop appropriate technological aids to assist Fontanella's transcriptions of the Robinson Reports, and to prepare for the development of the electronic search system. Templates and tags had to be developed before Fontanella could transcribe any reports. These templates will provide Fontanella with the basic structure of the transcriptions. The tags were implemented so that the information

from each report relevant to the electronic search system could easily be extracted. While Professor Fontanella is transcribing the reports, he will be using the tags that have been defined to mark up important information within the reports. To help him with this, a detailed user manual was developed that explains how he should use the tags and templates in conjunction with the XML editor.

Once the templates and tags were completed, a design strategy was developed for the completion of the system. The development of this design strategy required extensive collaboration with the NAL to determine the desired specifications of the electronic search system. The final specifications and recommendations were documented for use in the completion of the system.

This project report presents background information on cataloguing techniques and the different standards and languages: XML, TEI, SGML, and CIMI. It also provides valuable insight into who John Charles Robinson was and what he did for the V&A. Finally, the project report describes the methods that were employed by the project team, and presents the results and recommendations that were derived from those methods.

2 Literature Review

This literature review contains background information on several subjects of particular importance to the transcription of the Robinson reports. This research includes biographical information on Robinson and the significance of his letters, various cataloguing and archiving techniques, electronic cataloguing formats, techniques in information retrieval, and information on the Standard Generalized Markup Language (SGML) and the Text Encoding Initiative (TEI).

2.1 Sir John Charles Robinson

Sir John Charles Robinson was one of the most influential people in bringing the Victoria and Albert Museum to its present stature. Henry Cole, the original director of the museum, hired him in 1853 as the Superintendent of Art Collections. Robinson joined a governing body made up of the artist Richard Redgrave, the designer Owen Jones, and the German architect and theoretician Gottfried Semper (Baker 27). From the beginning of his tenure, Robinson made it his personal goal to make the V&A the greatest museum in the world (Baker 151).

Robinson's artistic knowledge was so diverse and authoritative that Cole, Redgrave, and the museum's other governing parties seldom refused his requests. This great knowledge of art was acquired at a young age, during his schooling at the studio of Michel Martin Drolling in Paris. While in Paris, Robinson was frequently found at the great French art museum, the Louvre. Visiting the Louvre stimulated Robinson's interest in Renaissance art (Robinson webpage).

After his years at the studio of Michel Marten Drolling, Robinson continued to study the arts. He became the headmaster of the Government School of Art at Hanley, Staffordshire, where he acquired his "lifelong passion for aesthetic and technical excellence in ceramics and sculpture" (Baker 151). Robinson remained at the school for six years until being offered a job by the V&A (Baker 151).

After joining the museum, Robinson began travelling throughout Europe, primarily in Spain and Italy. Robinson used his charm and great artistic knowledge to find and purchase works of art. If he was unable to persuade people to sell him the art, he often convinced them to allow photographs, drawings, or plaster casts to be taken of the items. In doing so, Robinson was able to build one of the greatest decorative art collections the world has ever seen (Baker 154).

Taking advantage of Italy's political troubles in the 1850s and 1860s, Robinson was able to construct one of the world's greatest collections of Italian Renaissance sculpture, majolica, metalwork, textiles, and furniture. He used his charm and distinguished contacts to convince people to sell him their valuable art. Many noblemen at the time were in dire need of money; their political stature was diminishing and to continue to live their luxurious lives they were forced to sell off their possessions. Robinson took full advantage of their political troubles. Toward the end of his career he was quoted as saying, "The noble masterpieces of sculpture now at South Kensington... [A]s a collective series [they] have no parallel, even in Italy" (Baker 152).

In an attempt to sustain the original goal of the museum – to educate the general public about the arts – Cole and Robinson created a system of circulating artwork with

area organisations. This system provided people from other areas of England with a way of experiencing the V&A without having to travel to the museum (Baker 151).

During Robinson's tenure at the museum, his relationship with Cole and Redgrave steadily deteriorated (Baker 157). Redgrave especially critiqued Robinson for his need to buy original artworks; Redgrave felt that it would be more cost-effective and serve the same purpose to have the pieces copied. Robinson defended his purchasing of originals, saying that copies lose the essence of the piece (Baker 152). Eventually, his differences with Cole and Redgrave became too much to overcome, and Robinson's power within the museum began to decline. The first step in this decline was his demotion from Superintendent of Art Collections to Permanent Art Referee in 1863. He remained at this position until Cole forced him out of the museum completely in 1868 (Baker 157).

Even after being driven out of the museum, his intuition for great art never ceased. He was still considered by many to be one of the premiere art connoisseurs of the world. He went on to become the Surveyor of the Queen's Pictures from 1882-1901, and was knighted in 1887. He also gathered an impressive personal collection containing several notable pieces (Baker 158).

Robinson's impact on the museum was felt for years to come. After Cole's retirement in 1873, the museum strayed away from Cole's contemporary purchasing style to a more historical style. The museum started to purchase items because of the period they reflected or the style they depicted (Baker 159).

2.2 The Robinson Reports

During his journeys throughout Europe, Robinson was in constant contact with the museum, writing letters every few days. When he found a good price on a piece or collection, he would write to the trustees of the museum requesting funds to purchase it (Fontanella 25/1/00). He often made these requests in small lots, listing and describing several pieces that he had found, as well as stating the pieces' prices. The museum board would use this information to decide if the item should be purchased (Robinson Reports). He would also write to the trustees of the museum to supply them with inventories of artefacts he had already bought, to give descriptions of possible future purchases, to inform the museum of difficulties he had in shipping, and to keep the museum updated on where he was and what he was doing (Fontanella 25/1/00). Some of the letters are unique and do not fit into any of these categories; for example, one particularly long letter was written to the Lord President of the Council. This letter described several pieces, now residing in the museum, which Robinson had purchased and now believed to be counterfeits (Robinson Reports).

Robinson's letters were often signed or initialled by the people who had read them, including Henry Cole and Richard Redgrave. Quite regularly the readers would write answers to questions and other comments on the letters, either to each other or in response to Robinson. When Robinson was requesting to buy items, readers would often note which items were to be purchased (Robinson Reports).

As Robinson's tenure increased so did the number of letters, until several thousand had accumulated. To this day, Robinson's letters, along with those of subsequent Art Referees into the 1880s, remain at the National Art Library (NAL) in the

V&A. This collection is known as the Robinson Reports and contains all of Robinson's letters as well as responses, telegraphs, printed catalogues and other related materials. The Robinson Reports are regarded as very valuable and are safeguarded with extreme care by the NAL.

The NAL only allows access to the Robinson Reports to scholars with special permission to use them for research. The Robinson Reports are obviously rich with valuable information on the provenance of the artefacts Robinson purchased and the people he dealt with; the descriptions Robinson wrote about the pieces could give scholars new insights into their origins. In addition, the knowledge of where he bought items could prove valuable in tracing an item's history. The letters are also useful to scholars researching people that Robinson encountered; in some letters he gave his impressions of people, or simply stated where they were and what they were doing at a particular time. Information like this is often very useful, but difficult to find.

One example of the Robinson Reports being used for this purpose can be seen in Lee Fontanella's book, *Charles Thurston Thompson and the Iberian Photographic Project*. Charles Thurston Thompson was a professional photographer who worked for the V&A during Robinson's time. He photographed works of art that the museum could not purchase. These works ranged from cathedrals to pieces in the private collections of royalty (Fontanella 27-39). Fontanella used Robinson's letters to establish where Thompson was at certain points in his career and why he was there. He was also able to determine why Robinson and Thompson chose the pieces that were photographed. Examples of the letters that Fontanella used in his research are available in Appendix B.

Unfortunately, the Robinson Reports are very difficult to use. They are stored only in chronological order, without any other sort of cataloguing. It is very difficult for scholars to find information on a specific subject. When Fontanella was researching Thurston Thompson, he had to painstakingly search through each letter from the time period that Thompson was active (Fontanella 25/1/2000). This problem is compounded by the difficulty of reading the Robinson Reports. There have been many changes in penmanship since the 1860s, and the writing style used in the letters is very different from writing styles used today. Moreover, the letters were often written very quickly, which makes them even less legible.

Fontanella will be transcribing the Robinson Reports in the summer of 2000. This will address the latter problem, as the electronic texts will be entirely legible. However, the letters will still be very difficult for scholars to search for specific information. For this reason, a cataloguing and searching system is desirable.

2.3 Cataloguing

Cataloguing is the act of producing an arranged list that describes the contents and order of a collection. This could be a collection of books, artwork, music, letters, or anything else that must be easily accessible (Wynar 1). Cataloguing has three main objectives: describing the contents of the collection, identifying their principal subjects, and devising a system to locate items within the collection.

Naturally, cataloguing is not always necessary. With small collections, the owner can either remember all of the collection's contents and their respective locations, or can quickly sift through the collection to find the desired items. However, as a collection grows, it becomes harder to remember what it contains. Locating specific items in a

large collection can become difficult as well. If a collection becomes too cumbersome, it is advisable to adopt some sort of cataloguing system (Wynar 1).

2.3.1 Library Cataloguing

Nearly all modern day libraries use catalogues to organise their collections. Most libraries follow very explicit national and, to a large extent, international cataloguing rules in sorting their collections. Standards have been developed for all of the major concerns of cataloguing.

In libraries, books are catalogued according to their bibliographic information, which includes author, title and publication information. Rules regarding the bibliographic description of books are available in the *Anglo-American Cataloguing Rules*, 2nd ed., 1988 revision, or the *AACR2R* (MARC 1-2). The *AACR2R* specifies the number and type of entries that a book should have. It clearly describes the content and form of these entries (Wynar 27-8).

Subject headings are generally offered in catalogues for patrons seeking information on a particular subject without any prior knowledge of relevant books or authors. The headings are a breakdown of the overall topic that a book covers; these take the form of keywords or phrases, under which the book is filed (Wynar 223). The subject headings that are assigned to a book are carefully chosen to be as simple, narrow, and consistent as possible (Wynar 230). To help with this task, librarians follow guidelines and use standard lists of headings.

Libraries use classification systems to make books easier to locate. A classification system defines the organisation of books on the shelves. It also provides a way for the catalogue to point to the locations of specific books. Libraries usually

employ one of the standard classification systems that have been developed over the years (Wynar 165).

After the appropriate classifying and cataloguing information about a book is determined, this information is placed into a catalogue under its author, title, and subject(s). This allows patrons to conduct their research using a variety of methods, depending on the information they have about the book they are seeking.

2.3.1.1 Subject Headings

Subject headings have two main objectives in a library catalogue. The first of these is to identify books that contain information on a specified topic; the second is to help patrons find books on similar topics (Wynar 229). A clear and carefully chosen list of subject headings will accomplish the first objective. The second objective is often carried out through cross-referencing. Cross-references consist of "see" and "see also" directives that point to other subject headings where the patron may find more information on the specified topic (Wynar 231).

When subject headings are poorly thought out, finding information in a subject-oriented catalogue can be quite difficult. This could be due to confusing or uncommon terminology, very broad or narrow subject headings, inconsistency in the use of terms to describe a subject, or confusing and unnatural cross-referencing (Rowland 249). To address this problem, most modern libraries look to standardisation.

In the pursuit of standardisation, many rules have been formulated for the creation of subject headings. These rules range from very general ideas to explicit lists of allowed headings. Carefully following standards to choose subject headings helps create a more controlled and efficient catalogue and reduce confusion among patrons (Rowland 247).

The most generic of these standards is a list of three concepts that librarians are taught to keep in mind when categorising a book. First, all books on one subject should be classified under the same subject heading. Second, the headings should attempt to use terms that would be familiar or common to the library's main audience. Third, the heading should not be broader or narrower than the topic covered in the book (Wynar 227-8).

While following these guidelines, a cataloguer may run into the problem of choosing between synonyms, or the situation in which no words exist to describe a book's topic. If there are several words for a book's topic, a cataloguer is instructed to use the most well known, the most common in the catalogues of other libraries, that with the least definitions, or the most likely to be the subject of other books (Wynar 230). If no word exists to describe the subject, the cataloguer is told to make a short phrase or sentence that amply describes the book (Wynar 230).

Rules also exist for dictating cross-referencing procedures. Cross-references, librarians are told, should point from larger to smaller topics, but not vice versa. They can also point back and forth between similar or supporting subject headings. Finally, cross-references can point from a term that is not used as a subject heading to the term where listings on that subject can be found (Wynar 231-2).

These simple rules for devising subject headings can make it easier for library patrons to find what they are looking for. Unfortunately, headings still do not make catalogues consistent and make no guarantee that things will be located where a patron expects (Rowland 246). As a result, almost all libraries use a standard list of subject headings, carefully devised and refined by experts in the field. A standard list of subject

headings is useful for ensuring that a search under a certain subject locates all books on that subject (Furrie 2). If standards are not used in a library's catalogue, patrons may not find the best books available on a subject. The two most common of these lists are the *Library of Congress Subject Headings* and the *Sears List* (Wynar 235). However, if a library is very small or highly specialised, these lists are often too general. In these cases, custom subject headings need to be generated.

2.3.1.2 Classification Systems

Classification systems govern the order of books in a library. Generally, these systems divide the library's collection by subject. The subjects are then subdivided until the topic of each book is adequately described. Grouping books by subject allows library patrons who have found one useful book to easily browse the area to find more books on that subject (Wynar 165).

A well-designed classification system must meet several criteria. First, it must include all possible subjects, including those covered by books already in the library and those that could be acquired in the future. This exhaustive coverage reduces the necessity for alteration of the classification system as the library's collection grows. Second, the system should be orderly, placing similar subjects together. This will increase the likelihood of a patron finding material by browsing the stacks. Third, the system must be expandable so that new subjects can easily be added (Wynar 179-80).

Once books are classified and shelved, a correlation must be made to the catalogue. This is done with a call number, which is usually indicated in the catalogue and on the books' spines (Wynar 165). A call number is an alphanumeric sequence that represents a specific subject area, and an area of the library where books on that subject

can be found (Wynar 204-6). Thus, if books are placed on the shelves in order of call number, they will be side by side with related books.

Librarians have developed several standard classification systems. Of these, the two most common are the Dewey Decimal System and the Library of Congress System. Both of these provide methods for categorising and organising libraries' collections, as well as clearly defined guidelines for developing call numbers (Wynar 181,207).

2.3.2 Archiving

Archives are records that represent an organisation's prior activities. These records are no longer needed for everyday use, but are kept and preserved for use in research (Miller 3). Because of their unique nature, archival records cannot use ready-made classification systems as libraries do (Miller 5). Instead, each collection must be individually arranged and catalogued. Archivists borrow heavily from the methods used in libraries, but alter them to meet their specialised needs.

The first goal in archiving is arranging records within a collection. To do this, the archivist groups together records with common provenance, then subdivides those groups according to the records' characteristics (e.g. form and content). Relationships are then distinguished between the various sets of records. These relationships govern the arrangement of sets of records within the collection (Miller 7).

After an archive is arranged, the records must be described and catalogued. Archival description is necessary for getting information about records and their authors to users. This description needs to be done in a well-organised fashion to make searching archives as easy as possible (Miller 79). Specific guidelines based on the *AACR2R* have been developed to aid archivists with description. These guidelines are laid out in detail

in Steven Hensen's *Archives, Personal Papers, and Manuscripts*, or *APPM* as it is known.

APPM covers two aspects of archival description: the bibliographic description itself, and the choice of headings and titles. Content and format for archival records are clearly described in *APPM*. Deviations from the *AACR2R* are especially stressed. *APPM* rules for headings and titles dictate the proper way of stating persons' names, geographical areas, titles of corporate bodies, and uniform titles[1]. These rules go into extreme detail and contain well over a hundred permutations of these categories.

2.3.3 Information Retrieval and Electronic Cataloguing

Information retrieval manages the problem of "locating relevant information from a body of widely dispersed knowledge" (Heaps 4). In today's world this is done mainly with the aid of computers. Computers are well suited for addressing the core issues of information retrieval, which are storing sufficient amounts of data and accessing that data rapidly. Although computers are well equipped to handle the problem posed by information retrieval, the programs they use must be well designed and easy to use.

To make information retrieval effective, computers must be able to interpret cataloguing records and search through large amounts of data efficiently. Standard formats have been created to allow computers to understand cataloguing records. Also, different methods for efficiently searching and retrieving data from a cataloguing system have been developed.

---

[1] Uniform titles are headings that are created by the cataloguer to bring like things together. For example, the use of the heading *Bible* to represent all the various translations of the Bible would be a uniform title (Hensen 133-4).

2.3.3.1 MARC Formats

Machine Readable Cataloguing (MARC) formats use a series of standard tags to allow computers to interpret a cataloguing record. MARC formatting has a distinct three-digit code, or tag, for each bit of bibliographic data, or field, in a record. A computer program can read these tags and, by replacing them with the appropriate text, reconstruct the original record. By using standard tags to represent each field, libraries can exchange records to reduce the amount of data entry that is necessary. They can also install pre-made electronic systems that read the records and automatically generate catalogues on computers (Furrie 1-6).

2.3.3.1.1 USMARC Formats

The standard MARC format for libraries in the United States is USMARC, developed by the Library of Congress. In USMARC format, the first digit of the tag represents the bibliographic area that a field belongs to; for instance, *1* represents the main entry. For a complete listing of the various definitions the first digit can hold, see Table 1 (Furrie 9).

Table 1 – Definitions of the first digit of MARC codes (Furrie 9)	
0	Control information
1	Main entry
2	Titles, editions, publication information, etc.
3	Physical descriptions
4	Series statement
5	Additional notes
6	Added entries filed by subject
7	Other added entries
8	Series added entries
9	Available for local definition

The last two digits of the tag represent a more specific area to which the field belongs. Some of these codes and their associated areas are 00 for a personal name; 10 for a corporate name; 30 for a uniform title; 40 for a bibliographic title; and 51 for a geographic name (Furrie 11). 9X is left open for local use (Furrie 9).

In USMARC format, there are two indicators after the tag; these indicators are single digits with a variety of uses. Depending on the field in which they are used, the indicators have different meanings. For instance, the second indicator in the title field is the number of characters the cataloguing program is to skip while placing the entry in alphabetical order. This is useful for making the program skip over initial articles in titles. If an indicator is not used, a pound sign (#) is put in its place (Furrie 8).

After the indicators in the USMARC record, there are subfield codes. These codes are used to break the field up into smaller pieces, and are represented by a delimiter (generally "$", but sometimes a different character such as "_" or "@") followed by a letter (Furrie 8-9). For example, $a and $c represent *title proper* and *statement of responsibility* when used within the title information field (Furrie 4-5).

A full USMARC entry for one field may look something like this:

260 ## $a Allentown, PA : $b Bill's Books, $c c1979.

The tag, 260, represents the publication information field. In this field, neither indicator is used. Hence, their places are filled with pound signs. The $a, $b, and $c are subfield codes that stand for place of publication, name of publisher, and date of publication. Cataloguing software would interpret this entry as follows:

Publication, Distribution, etc., area:

Place of Publication: Allentown, PA :

Name of Publisher: Bill's Books,

Date of Publication: c1979.

Significance of all of the USMARC tags, and their respective indicators and subfields, is available in *USMARC Format for Bibliographic Data*, published by the Library of Congress (Furrie 4-5).

2.3.3.1.2 USMARC AMC and ArtMARC

Some variations of the USMARC format have been developed for areas of cataloguing outside of libraries. Among these are USMARC AMC (Archival and Manuscripts Control) and ArtMARC.

The USMARC AMC system was developed to aid archivists in cataloguing. While conforming to USMARC standards, the AMC format alters and adds some fields to serve its purpose. One major feature of the AMC format is the ability to link records. This allows archivists to make one record for a collection, and link it to individual records for the contents of that collection (Miller 114).

The ArtMARC format takes this idea even further in altering USMARC to catalogue physical objects. The cataloguing of art requires entirely different information from that of books. Scholars and historians in the fields of art and architecture usually need a more extensive history of an artefact than librarians need of a book. For instance, they may want to know where it came from, who made it, and what it represents, along with a myriad of other details (McRae 5). The ArtMARC system was developed to accommodate this sort of information, while still maintaining compatibility with the USMARC standards.

2.3.3.2 Subject Indexing

An important characteristic of an electronic cataloguing system is providing the capacity to search through the catalogue efficiently. There are many methods for searching, such as a search by author or by title; however, the most popular search technique is the subject search. To allow the user to search through information by subject, a technique known as subject indexing must be implemented. There are two phases to implementing subject indexing. The first is determining "a set of words, phrases, or sentences that collectively represent the information content." The second is "deciding which of these are worth recording as being relevant to the interests of those who are expected to use the information system" (Vickory 76).

2.3.3.2.1 Automatic Indexing

When cataloguing a document, a machine can complete all the stages of subject indexing as long as a few conditions are met. In order for a document to be analysed by a machine, the text of the document must be in a form that the machine can process. The machine should also be able to distinguish individual words in the document. Once the previous two conditions are met, there should be a method for providing the machine with rules. These rules define types of words and phrases it should use as indices (Vickory 77).

There are two methods for defining these rules; both methods have been used in experiments on machine indexing. The first method is to formulate a list of keywords for the machine to find. The second method is to develop a list of words for the machine not to include as indices; these words might be articles, pronouns, conjunctions, etc. The second method will probably produce numerous potential indices, many of which will not

be relevant. One way to remove irrelevant words from the list is statistical indexing, developed by H. P. Luhn. Statistical indexing keeps track of the number of times each word occurs in the document. The words that appear the most are selected for indexing (Vickory 78).

2.3.3.2.2 Using Subject Indexing for Retrieval

Once the subject index is built, there must be a way to search through the index to locate the relevant information that the user is attempting to retrieve. This is achieved by developing a list of question words. The user, who is attempting the subject search, usually enters these question words. In order to retrieve the information the user is looking for, a match needs to be made between the question words and the subject indices. To see how both the index words and the question words affect performance, see Figure 1 (Vickory 80).

In Figure 1, AB represents the entire collection of documents. In a given document search, AC represents all of the data relevant to that search, whereas CB represents all of the irrelevant data. A search, however, rarely produces the set of

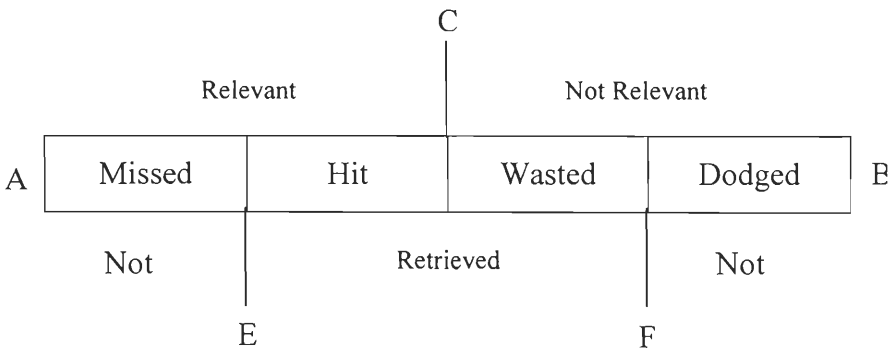

Figure 1 – Search and Retrieval Model (Vickory 80)

documents AC; it usually produces a set of documents EF, some of which are relevant (EC) and some of which are not (CF). The diagram also shows that some relevant documents may be missed. In order to improve the retrieval system, the search results must be realigned to coincide as closely as possible with the set AC.

The simplest way to accomplish this is to choose as many index words from the document as are needed to describe it. There are three factors to consider when doing this: the number of index words per document, the control of indexing the words, and the number of access points per document (Vickory 80-86). To find out more about these factors and about subject indexing in general, see B. C. Vickory's book *Techniques of Information Retrieval.*

2.4 Standard Generalized Markup Language (SGML)

The Standard Generalized Markup Language, known as SGML, is an international standard for the definition of device-independent, system-independent methods of representing documents in electronic form. SGML is a metalanguage, which is a definition or description of language. Specifically, SGML defines a method for describing a markup language (TEI Guidelines Part I Sec. 2).

The term *markup* refers to any means of specifying the interpretation of the text within the text itself. This is done to a limited extent in plain text. Having a group of words logically separated with punctuation is a form of markup. For example, an exclamation mark is placed at the end of a sentence to show emphasis; this low level form of markup makes it easier for a reader to interpret the text. SGML is a more

complex type of markup that provides a standard way in which computer applications can interpret electronic text (TEI Guidelines Part I Sec. 2).

A markup language refers to a collection of markup rules that are used to encode text. The language must define what markup can be used, what markup must be used, how the markup is differentiated from the text, and the meaning of the markup. SGML fulfils all of these requirements except for defining the meaning of the markup, for which documentation must be provided (TEI Guidelines Part I Sec. 2).

2.4.1 SGML Document Structures (DTD)

SGML fulfils the first two requirements of a markup language through the use of a *document type definition*, or DTD. A DTD is the "formal specification for the structure of an SGML document" (TEI Guidelines Part I Sec. 2.4). Each DTD is a specific interpretation of a text. Because there are different factors that must be addressed when choosing or defining a DTD, there is no ideal DTD for a specific type of text. The type of analysis that will be done on the text will greatly affect the chosen or defined markup that can and must be used. Therefore, it is very important to consider the type of analysis when choosing a DTD. For example, if a DTD for marking up the text of short stories was being defined and the topic of interest was characters, character names would be of interest and would be marked up. If the same short story needed to be marked up but the topic of interest was the structure of the text, there would be no need to mark up character names; in this case, it would be more important to mark up paragraphs, pages, and indentations (TEI Guidelines Part I Sec. 2.4).

There are certain components that make up a DTD. The two basic components are elements, with their associated attributes, and entities.

2.4.1.1 Elements and Attributes

To identify a textual unit, such as a paragraph or a phrase, SGML defines the concept of an *element*. SGML allows each element to be uniquely identified by a specific name, which is defined in the DTD. However, SGML does not provide a way to define the meaning of the element because the SGML standard is not concerned with semantics. It is up to the designer to give meaning to the elements through documentation or other means (TEI Guidelines Part I Sec. 2.3).

In the SGML standard, elements are allowed to have *attributes* associated with them. In a technical sense, an attribute allows for the incorporation of additional information that describes a particular instance of an element, without that information being part of the content of the element (TEI Guidelines Part I Sec. 2.6). For example, it might be desirable to have an "id" attribute associated with a letter element in order to reference that element from another point in the text. The information stored in the attribute is not part of the content of the letter, but is associated with the letter element and can be used to distinguish a particular letter element from another letter element. In SGML, this is done in the following manner:

```
<letter id=''letter1''> … </letter>
```

In this example, other letters could refer to the letter element as "letter1." For more information on referencing elements from other locations in the text see *TEI Lite: An Introduction to Text Encoding for Interchange (Section 8)*.

2.4.1.2 Entities

Whereas elements as used to mark up structural aspects of a document, *entities* are used to mark up random parts of the document's content. For example, an entity

named SGML could be defined, where SGML's value could be the string "Standard Generalized Markup Language." To include an entity in an SGML document, a construct known as an *entity reference* is used. Whenever an SGML parser, a piece of software that can process an SGML document, encounters an entity reference, it replaces the reference with its associated value. There are many benefits to using entities; one such benefit is that it reduces the amount of typing necessary. For example, if a company has a standard document where its address always appears on the document, instead of always requiring someone to manually type in the address, an entity reference whose value is the address could be inserted in the text. Entities also provide another benefit that can be demonstrated using the last example. If the company's address changes, the address can be modified in all of the documents without actually changing any text by providing a new value for the address entity (TEI Guidelines Part I Sec. 2.7). For more information on what can be done with entities refer to the *TEI Guidelines for Electronic Text Encoding and Interchange* (Part 1 Section 2.7).

2.4.2 SGML Document

Once a DTD is defined for a specific type of text, an SGML document can be created using that DTD. An SGML document consists of two parts: an *SGML prologue* and a *document instance*.

The SGML prologue contains an SGML declaration and a document type definition. The SGML declaration is used to specify information about the dialect of SGML being used. It would contain information such as the codes used for SGML delimiters and the character set. This part of the declaration is usually invisible to the user; the SGML software being used to create the document usually generates it. The

document type definition part of the prologue is where the document is defined to be associated with a particular DTD. In this section of the prologue it is possible to declare additional entities and elements, or even to redefine ones that are defined in the DTD (TEI Guidelines Part I Sec. 2.9).

The document instance is made up of the content of the document being encoded and the markup used to encode it, as well as any entity references (TEI Guidelines Part I Sec. 2.9.3). Here is an example of a document instance:

```
<letter>
<letterHead> ... </letterHead>
<Salutation> Dear Miss Smith </Salutation>
      <body>
      Here is some sample text.
      </body>
<Closing>
      <Adieu> Yours sincerely </Adieu>
      <Author> Alexander Day </Author>
</Closing>
</letter>
```

Notice that the SGML convention for explicitly marking or tagging each element is to enclose the text or other elements, which make up the element, with a start tag *<element name>* and an end tag *</element name>* (TEI Guidelines Part I Sec. 2.3.1). For more information on the method of tagging elements and specifying attributes for elements, refer to the *TEI Guidelines for Electronic Text Encoding and Interchange (P3)*.

2.4.3 Advantages of SGML

SGML provides three main advantages over other markup languages; it defines a descriptive markup language rather than a procedural markup language, allows for the definition of document types, and is a non-proprietary format.

2.4.3.1 Descriptive Markup Language

SGML is used for defining markup codes, which describe certain parts of a document. For example, a person's name found in a text might be tagged with the <signed> markup code to signify that this text represents a signature of the person. In contrast, a procedural markup language specifies what to do with a particular section of the document. For example, a markup code <indent 0.5> would be used to signify that a particular portion of text must be indented by 0.5 inches (TEI Guidelines Part I Sec. 2.1.1).

By separating the formatting from the markup, SGML allows different pieces of software to format the elements that are marked up in different ways. For example, one piece of software might display all elements marked as a signature in a special script font to make the text appear as if it were a real signature. Another piece of software might display the element in normal font, but append a little note next to the text signifying that it is a signature.

The fact that elements are marked descriptively is also very important for programs that analyse the text to extract meaningful information from it. For example, a search program could locate documents that contain a certain person's name. If the documents were marked up using a DTD that allowed for the tagging of people's names, the program would only need to extract all the elements that were tagged as people's names and search the names for the person of interest. If the program found the person's name, it would know that the current document that it was searching contains a reference to the person.

2.4.3.2 Document Type Definition

SGML provides the concept of a *document type*. The type of the document is defined in a document type definition, or DTD, which offers several valuable advantages. All documents claiming to be of a certain type can be checked against their corresponding DTD. This allows programs to validate SGML documents to make sure they are of the proper structure. Another advantage gained by having document types is that all documents of the same type can be processed in the same way, which allows programs to handle documents more intelligibly (TEI Guidelines Part I Sec. 2.1.2).

2.4.3.3 Portability

Valid SGML documents are not dependent on a certain hardware or software environment; thus they are easily transported from one environment to another without the loss of any information (TEI Guidelines Part I Sec. 2.1.3). This allows the same SGML document to be read on any computer and by any software that can read SGML.

2.4.4 Extensible Markup Language (XML)

The *extensible markup language*, or XML, is a special markup language that provides a standard way to "identify and describe data on the web" (Oracle about_xml.htm). XML was developed in response to the limitations of HTML and the complexity of SGML (JavaWorld jw-04-xml_p.htm).

2.4.4.1 HTML Limitations

HTML, one of the languages defined by SGML, was designed to describe documents for the web, known as web pages. HTML specifies characteristics such as

headings, titles, and fonts used in the document. The limitations of HTML are that it is not extensible, is "display-centric," is not directly reusable, provides only one view of the data, and has little or no semantic structure (JavaWorld jw-04-xml_p.htm). Each of these limitations will be described in greater detail in the following sections.

2.4.4.1.1 Non-Extensible

An extensible markup language allows for the development of custom tags specific to particular applications (JavaWorld jw-04-xml_p.htm). This is not the case with HTML; there must be a pre-defined set of tags used for specifying the layout of web pages. Web browsers must be able to recognise these tags in order to display the document properly. It is not feasible to attempt to have all web browsers support all application-specific tags created by web developers. Therefore, web developers are forced to work with the standard set of tags provided by the World Wide Web Consortium (W3C) (JavaWorld jw-04-xml_p.htm).

2.4.4.1.2 Display-Centric

The purpose of HTML is to define the layout of information for a web browser to display; it allows users to tag information for display purposes. For example, text can be tagged as bold or italic. The problem with this approach is that such tags are meaningless to network applications, which focus on the meaning behind the data and not how it is presented (JavaWorld jw-04-xml_p.htm).

2.4.4.1.3 Reusability

HTML is not a language that is directly reusable. For example, if a web page is a representation of a word-processing document and the data in that document changes

after the web page has been generated, then the HTML for the web page must be entirely regenerated. In HTML, there is no way to separate the content and structure of the data from its presentation. If such a separation were possible, the presentation of the data would only have to be reapplied to the new data without the regeneration of the entire web page (JavaWorld jw-04-xml_p.htm).

2.4.4.1.4 One "View" of the Data

HTML does not make it easy to display a given document in different formats tailored to users or their browser's capabilities (JavaWorld jw-04-xml_p.htm). To address this need, a new type of HTML has been developed called Dynamic HTML. However, even Dynamic HTML has some major drawbacks, such as extensive scripting (JavaWorld jw-04-xml_p.htm).

2.4.4.1.5 Semantic Structure

HTML represents data by layout. It would be more beneficial for web applications if the data were represented by meaning. This would allow for vast improvements in searching over the web. If data were represented by meaning, then a search engine would be more efficient in finding information over the web (JavaWorld jw-04-xml_p.htm).

2.4.4.2 Uses of XML

XML was developed to overcome all of the shortcomings inherent in HTML. XML provides a whole set of new possibilities for web applications. XML is often used in web applications to customise the manner in which data is presented. It is also used to transfer data among applications (Oracle about_xml.htm).

2.4.4.2.1 Customising the Presentation of Data

Due to the fact that XML defines the structure and content of data rather than its layout, the way in which the data is presented is a separate concern. The presentation of XML data is defined through *XSL stylesheets*, which are text documents that specify a way to transform the XML into HTML or some other text-based format. The text documents are specified using XSL, the *extensible stylesheet language*. This method allows for the presentation of data defined using XML to be displayed differently for different devices, users, or browsers. Each manner in which the data are to be displayed simply must have its own stylesheet. This allows applications to deal with the operations they need to perform, without concern for the presentation of the data (Oracle about_xml.htm).

2.4.4.2.2 Exchanging Business Data Among Applications

Application developers commonly have problems when comparing and exchanging data among applications of different vendors. XML is a way of exchanging data that allows applications to focus on the format of the data without requiring the use of a specific application or protocol. Therefore, with XML and XSL stylesheets, information can be transferred between applications in a format that can be read by other applications (Oracle about_xml.htm).

2.4.5 Text Encoding Initiative (TEI)

The *Text Encoding Initiative* is an international organisation of researchers striving to create a standard means of representing textual materials in electronic form (Burnard Sec. 2). The TEI project began with two main goals: to decide which elements

of the text should be encoded in an electronic format, and to develop a method of encoding the text that will provide for platform-independent interchange without any loss of content or form (Burnard Sec. 2).

The Standard Generalized Markup Language was chosen to define the guidelines that would be used to represent the textual materials in electronic form. Using SGML, TEI produced over 400 SGML elements with their associated attributes to be used to mark up the textual material. TEI then grouped the elements into different tag sets. These tag sets can be combined in different ways to encode all types of text in any language and from any time period in a manner independent of hardware, software, and application (Burnard Sec. 2).

2.4.5.1 TEI Schema

As mentioned above, TEI has combined its elements into tag sets; these tag sets are what make up the TEI schema. A *tag set* is a group of SGML elements and their attributes. All elements in a tag set share some logical relationship with each other (Burnard Sec. 3). For example, one tag set could be the collection of tags that are used when marking up poetry. The tag sets are divided into four groups:

- **Core** tag sets – These tag sets define elements that appear commonly in all types of documents. They must be included in every TEI-compliant DTD.
- **Base** tag sets - These tag sets define elements that are particular to certain types of documents. Each TEI-compliant DTD should contain only one base tag set.
- **Additional** tag sets – These tag sets include elements that pertain to a specialised application or a specific subject area. These elements are not associated to a

specific document type or base tag set, but instead can be combined with any base tag set.

- **Auxiliary** tag sets – The elements defined within these tag sets are highly specialised. They are used for several reasons, such as describing a certain part of the encoding scheme. These elements are not associated with the main DTD, but rather make up their own DTD (Burnard Sec. 3).

These tag sets are combined to provide a specific "view" of the TEI DTD, which is specific to the certain type of document being defined. These tag sets are combined in the following manner:

- One base tag set (chosen)
- One core tag set (always present in schema)
- Zero or more additional/auxiliary tag sets (chosen)

There must always be both a base tag set and a core tag set in the DTD. However, the use of one or more additional and/or auxiliary tag sets is optional and in the discretion of the document designer (Burnard Sec. 3).

2.4.5.2 Consortium for the Computer Interchange of Museum Information

The Consortium for the Computer Interchange of Museum Information, also known as CIMI, is an association that customised the TEI standard for exchanging museum information. CIMI formed in 1990 with the intent of "bringing museum information to the largest possible audience" (CIMI home page Introduction).

CIMI achieved its goal by setting a standard for the transfer of information between museums. After years of research, in 1993 CIMI made the decision to support two standards; SGML for information structure and ANSI Z39.50 for information search

and retrieval (CIMI home page History Section 1). In order to incorporate these standards CIMI developed Project CHIO, or Cultural Heritage Information Online. This project was subdivided into two separate parts: CHIO Structure and CHIO Access (CIMI home page History Section 2).

CHIO Structure was completed first. The goal of this part of the project was to design a standard way to mark up documents containing museum information; the TEI tag set was used as a starting point. TEI was established exclusively to assist scholars. CIMI adapted the TEI tag set by adding custom tags that pertained to their goal of transferring information between museums. The final outcome of the CHIO Structure project was the CIMI DTD, along with its tagging guide (CIMI home page History Section 2).

The second part of the project, CHIO Access, used the CIMI DTD and tagging guide to develop an application of the Z39.50 standard, which is the ANSI standard for search and retrieval. In order to develop an application of the Z39.50 standard specifically for museum content, an "Application Profile" was defined. This profile holds the definitions for the type of queries that will be used to extract information from marked up documents, thus facilitating search and retrieval (CIMI home page History Section 3).

3 Methodology

We completed several tasks to aid Professor Fontanella with the transcriptions of the Robinson Reports and to prepare for the development of an electronic search system. We began by researching Robinson and his letters, relevant cataloguing techniques, electronic systems, and the Standard Generalized Markup Language. This research, as presented in the previous section, was applied to the various aspects of our work. Building from this background research, we began working toward achieving our goals. This work included developing and testing a tag set, transcription templates, and a user manual for transcribing the reports. We also completed several sample transcriptions, an estimate of the amount of time it would take to transcribe the collection, and a set of guidelines for the team that develops the search system. However, the first task that we needed to complete was to develop feasible goals and objectives for our work.

3.1 Development of Goals and Objectives

Before we could proceed with our project, we needed to clearly define our goals and objectives. In order to accomplish this we needed to develop a working knowledge of what the Robinson Reports are and what type of system we were expected to build. The first step in attaining this knowledge was to interview Professor Fontanella. Because we knew very little about Robinson and his letters, we staged this initial interview as a discussion, rather than a question and answer session. This allowed Professor Fontanella to control the conversation and provide us with information we would not have known to ask for. This interview gave us a better understanding of the Robinson Reports and the

system we were asked to develop. To guide our discussion with Fontanella, we developed several open-ended questions, which are available in Appendix C.

Our original goal statement was based on the information we gathered from our initial meeting with Fontanella and the description of the problem that the NAL had provided. The initial goal of this project was to design and build an electronic search system for the Robinson reports. Design criteria included being able to handle data in a non-proprietary format, such as XML, and display it to the user, preferably on the web. This system was meant to hold important information from the letters, but not the full texts. It would be used to find and identify specific letters so that they could be looked up in the transcriptions.

Our preliminary research revealed that this goal was not achievable. We had neither the time nor the technical background required to design a system that handles data in a non-proprietary format. We met with Professor Nabil Hachem of WPI, who is an expert on database development, and Ken Dobson, a professional in the field who has constructed systems like this in the past, to discuss software packages and implementation techniques that could be used in creating this system. We discussed the nature of different types of database software and the expertise and time required to implement them. We discovered that it would take a minimum of three months to construct this system if we already had the necessary expertise. Because we did not, it would take even longer to complete. Furthermore, if we were to complete this project as originally proposed, it would be too limited in scope to be an Interactive Qualifying Project (IQP); we would further develop our technological skills, but the project would be focused almost entirely on technology rather than the humanistic and societal issues such

as the needs of transcribers and of art historians. For these reasons, we needed to re-examine the problem and create new goals and objectives that would be feasible, helpful to the NAL and to Fontanella, and within the definition of an IQP.

We were able to redefine our goals by working closely with Fontanella to determine his needs and by examining our capabilities to determine what we could accomplish in the time allotted. We discovered that Fontanella had not fully decided how he was going to transcribe the letters. He did not yet know what application he would use to edit the transcriptions or what format he would be storing them in. Therefore, one of our goals was to prepare for, and to simplify as much as possible, Fontanella's transcription of the Robinson Reports.

Fontanella agreed to tag important information, such as names and locations, as he transcribes the letters. We had planned to develop an entirely customised tagging system that he could use to mark relevant information, which could later be entered into a separate cataloguing system. However, after talking with Douglas Dodds, the computer specialist at the NAL, we realised that it would be much more beneficial for all parties to use a standard set of tags, modified to fit the unique nature of the Robinson Reports, that would be recognised by organisations around the world. We also found that it would be better to incorporate the entire texts of the transcriptions into a searchable system, as opposed to creating a separate cataloguing system. Hence, our second goal was to prepare for the development of an electronic search system that stores the full text of the Robinson Reports in a standard non-proprietary format.

We developed several objectives that we needed to meet in order to accomplish these goals. To aid Fontanella with the transcriptions, we researched and chose an editor

for him to use when transcribing the Robinson Reports. In addition, we developed templates to specify the structure and style of the various types of reports and reduce the amount of time that Fontanella would need to spend on formatting. We also created shortcuts to the most common phrases in the reports so that he will not need to tediously re-enter phrases that show up very frequently. Finally, we made an estimate of how long the transcriptions would take and how many Fontanella will be able to do in the time he is working at the NAL.

To prepare for the development of the electronic search system, we defined a standard method of marking up the transcriptions. Furthermore, we developed a system for tagging important information in the transcriptions. After we completed these objectives the system will be able to easily search through the transcriptions and extract relevant information.

3.2 Interviews of Cataloguers

As we were developing our goals and objectives, we conducted many interviews to discover how electronic cataloguing systems have been implemented for other collections. For these interviews we contacted Lora Brueck in Special Collections Management at the WPI Library, Jill Burns of the Worcester Art Museum, and Alan Degutis of the American Antiquarian Society. We used interviews because they provide more up to date information than published materials, as well as an interactive method of gathering it. This interaction was important because it allowed us, as the interviewers, to control the flow of information, whereas published materials grant control to the expert. In addition, conducting face-to-face interviews provided us with an opportunity to see the cataloguing systems, which helped us gain a better understanding of the system that is

required for the Robinson Reports. We asked these experts direct open-ended questions regarding cataloguing methods, system architecture, and software choice. These questions are available in Appendix C.

We found that the WPI Library and the Worcester Art Museum both use commercial cataloguing systems that provide intuitive graphical user interfaces. However, neither handles data in a non-proprietary format. The system at the American Antiquarian Society was in a non-proprietary format, but had a complicated text-based interface. The system that we sought to develop would be a combination of these two. Seeing these systems at these organisations reinforced the notion that it would be implausible for us to design and implement the entire system in just seven weeks. However, the experience also helped us to better understand how systems of this nature are used in organisations; this knowledge helped us to make better and more beneficial recommendations to the final development team.

3.3 Selection of Editing Software for Transcribing the Reports

One of the first things we had to do upon arriving in London was to choose an editor with which Professor Fontanella would transcribe the Robinson Reports. There were three criteria that this editor had to meet. First, Fontanella had to be comfortable using it. Second, it had to support an SGML/XML editor. Finally, it had to provide the capabilities of building templates and defining shortcuts.

Fontanella must be comfortable using the editor because he does not have an extensive technical background. More importantly, he will be pressed for time when transcribing the Robinson Reports. He will be in London for only seventy-five days;

therefore, he will not be able to spend a lot of time learning how to use a complicated and unfamiliar editor.

We needed to choose an editor that supported SGML and XML because the NAL wants the transcriptions to be stored in a non-proprietary format. SGML and XML were chosen because they have become the standard for storing data in a non-proprietary manner. Additionally, SGML and XML support the implementation of custom tags, which would be advantageous for a collection as unique as the Robinson Reports. Moreover, SGML and XML allow for importing and modifying standard tag lists that are commonly used for marking up manuscripts and documents. For more information on SGML and XML see Section 2.4.

Finally, the editor must provide the capabilities to create templates and define shortcuts. As decided in our interviews with Fontanella, we would be creating templates to facilitate his transcription of the Robinson reports. The editor we chose had to allow us to define the basic format of the reports to provide Fontanella with a consistent structure for transcribing the reports. Additionally, we wanted to provide shortcuts to common phrases in the reports to save Fontanella typing time. In order for us to provide this feature, the editing software had to allow for the definition of custom shortcuts.

With these three criteria in mind, as well as the limited funds for the project, we began doing research on the web. Web research was necessary because the formats and standards we used are still under development; hence, there is a lack of published information available on these technologies. While conducting our research, we read various reviews and compared the advantages and disadvantages of many available editors. Based on our findings we chose an editor and requested funds to purchase it.

3.4 Analysis of the Robinson Reports

Concurrently with our research into editors, we began to explore the Robinson Reports. First, we looked over a sample of letters that covered the span of the entire collection. This gave us an idea of their general appearance and content. It also helped us determine the size of the collection, which we used to pace ourselves in our work.

After our initial perusal of the reports, we read through the entire collection and carefully examined each of the reports. We analysed the content and form of the Robinson Reports, searching for unique formats and styles, important information that may need to be tagged, and common phrases.

3.4.1 Unique Formats and Styles

In order to construct the templates, we needed to identify all of the different formats and styles represented in the Robinson Reports. Different types of documents, such as detailed shipping records, minutes, and telegrams, would each have their own template. We carefully went through the boxes and recorded the box and report numbers of any distinctive styles and formats we could identify. After sifting through the first few boxes, we had recognised most of the major formats. We then continued searching through the remaining reports to locate any reports that had a different format from those that were already recorded.

3.4.2 Information to be Tagged

While reading the Robinson Reports we developed a list of information that could be tagged in the transcriptions. These tags will serve two main purposes. Many of them will be used to identify information that a scholar would use to search for particular

transcriptions; for instance, artists' names would be tagged so that a scholar could search the transcriptions for the mention of a particular artist. The remainder of the tags would be used solely for altering the display of the text. For instance, if "J.C. Robinson" was tagged as a signature, the program that displays the text could interpret it as such. It could then display it in italics, with (signature) appended to it, or in any other manner.

The development of tags required an extensive review of the reports. Before conducting our research, we were given several tags that Professor Fontanella had already established as necessary: chronology and origin of the report, classification and provenance of artefacts mentioned, and the geographical areas where they originated and were purchased.

We also spoke with Christopher Marsden, the assistant archivist at the NAL, to help us better understand how to categorise the tags that identify specific reports. Several identifying numbers have been written on the reports. Mr. Marsden explained to us what each of these numbers meant, what they were called, and which were relevant to the archives today. This helped us decide which numbers would require tags and which could be ignored.

Using the ideas that we gathered from Fontanella and Marsden as a guide, we began research into other information that could be tagged. While reviewing the letters to identify their various formats, we also examined their content extensively in order to develop relevant tags. Our previous research into cataloguing and archiving techniques was very useful to us in this endeavour, as it gave us a better understanding of the type of information that would be useful. Our knowledge of subject headings and the purpose of archives was particularly useful in helping us choose relevant and practical tags.

As we progressed through the letters, we constructed a list of all types of information and formatting that might be tagged. While constructing this list we remained very open-minded, never ruling out any information that scholars might possibly want tagged.

3.4.3 Common Phrases

In addition to identifying the various formats of the Robinson Reports and developing a preliminary list of tags, we took note of any phrases that occur regularly throughout the collection. For instance, many of the letters refer to the South Kensington Museum, as the V&A was called during Robinson's time. We kept track of these phrases because they could be defined as shortcuts in the SGML/XML editor. The editor will then allow the user to insert them either by using their assigned keystrokes or by pressing a button on the editor's toolbar. This would save Professor Fontanella a lot of time, as he would not have to type frequently-used phrases over and over again.

We used two simple criteria in choosing common phrases to define as shortcuts. First, and most obviously, the phrase had to occur regularly throughout the collection. Second, it had to be long enough to save a significant amount of typing time. For example, "H.C." (the initials of Henry Cole) is on almost every letter; however, it would be just as simple for Fontanella to type "H.C." as it would be for him to access a shortcut. We also tried to keep the number of phrases low; if we defined too many shortcuts, Fontanella would be hindered by a multitude of keystrokes and cumbersome toolbars. It would be simpler for Fontanella to type "South Kensington Museum" than to sort through hundreds of buttons or complex keystrokes to insert it.

3.5 The Development of Tags

After completing our analysis of the Robinson Reports, we were ready to create the document type definition (DTD) for the transcriptions. A DTD defines the rules that specify the structure of an XML document. (See Section 2.4.1 for more information.) Defining the DTD was a complicated process that required a lot of research and thorough analysis and development. First, we compared the list of tags that we drew from the Robinson Reports with the TEI and CIMI standards. We were then able to finalise the tag list and define a DTD containing the relevant parts of the TEI tag set and the necessary CIMI tags.

3.5.1 Comparing the Initial Tag List with the TEI and CIMI Standards

After all the possible tags were recorded, we carefully reviewed the standard lists of tags included in TEI and CIMI. We compared the tags in our list to those tags, looking for any matches we could find. This required an open-minded comparison between our abstract ideas and carefully interpreted definitions of the standard tags. This was especially difficult due to the deficiency of documentation for the TEI tags. The TEI standard is still under development and not yet fully documented for public use. During our comparison, we noted which tags were covered by standards, which tags could be fit into the standard, and which tags were not covered at all.

Many of our tags had existing counterparts in the standards. These tags were recorded, along with the corresponding TEI or CIMI tag and a brief explanation of the tag's purpose. When deciding between multiple standard tags, we tried to avoid using the CIMI standard in favour of TEI. This was done because of the long and difficult process involved in adding new tags to the TEI DTD.

Some tags could be fit into the standard, but did not have exact counterparts or could not have all of their attributes represented in the standard. These tags were marked as such. For instance, the box and page numbers of a given report can be stored in the "id" and "n" attributes of the TEI header tag. This would work, as long as the team developing the search system was made aware of it, but it is not particularly intuitive or standard. Another example of a tag that could fit into the standard, but not very well, is the date tag. There is a tag called "date" in the TEI standard, but the Robinson Reports contain several different types of dates. The TEI tag does not have a "type" attribute, so there is no way to differentiate between different kinds of dates.

Finally, a small portion of the tags we developed had no counterparts at all in the standards. These tags would either need to be eliminated, or custom tags would need to be defined for them. We wanted to define as few custom tags as possible so that we could stay close to the standard and, as with the CIMI tags, avoid the difficult process of adding new tags to the DTD. However, we did not want to remove any tags that would be necessary or highly desirable for scholarly research.

The comparison we made between our tags and the standards helped us determine a final tag list, as we now knew which tags were already represented and which would need to be defined. It also helped us to decide which subsets of TEI to include in our DTD. The process of comparing each of our tags to the standards made it clear to us which tag sets would be necessary for the transcription of the Robinson Reports because we could see where our tags were located. For more information on the different types of tag sets within the TEI standard, see Section 2.4.5.1.

3.5.2 Finalising the Tag List

Once we had determined which of our tags were represented in the standard, we discussed the relevance of each with Professor Fontanella and members of the NAL staff: the head of public relations, John Meriton; the computer specialist, Douglas Dodds; and the assistant archivist, Christopher Marsden. They shared their knowledge and experience of what information is important to scholars and what information should be tagged for archival and bibliographical reasons. Therefore, they could better judge which tags should be added, changed, combined, or eliminated.

In these discussions, we explained that it was important to stay as close to the TEI and CIMI standard tag sets as we could. It would be acceptable to add custom tags, if they were for important and relevant information for which tags were required; however, adding unnecessary tags that were not a part of the standard would make the format more proprietary, which is highly undesirable. We also made it clear that CIMI tags should only be chosen when it was absolutely necessary, as this would save us time in implementing the final DTD. Finally, we explained that Fontanella would not necessarily need to insert all of the tags himself. Rather, if it was too much work to fit into his time frame, he could decide which tags he would have time to insert and someone else could insert the rest of the tags at a later date. This kept us from ruling out tags simply because he would not have the time to enter them.

During our discussions with the NAL staff, we considered the value of our tags and eliminated those that were not necessary. We suggested to them which tags we thought were important and how we intended to implement them; they gave us feedback regarding the acceptability and usefulness of those ideas. Mr. Dodds was very helpful in

suggesting technical solutions, as he has had experience with TEI and other forms of XML. Mr. Meriton, from the point of view of a scholar without a technical background, provided us with more abstract ideas that we could build on. Mr. Marsden helped us determine which tags to use for archival purposes, such as identification numbers. Through this interaction we were able to eliminate or find a standard tag for every one of our tags; therefore, we would not need to define any custom tags.

Our discussions with Fontanella, held after those with NAL staff, were used to determine his needs and desires. We presented him with the list of tags that we developed at the NAL and asked if there were any changes he would like to see made. Specifically, we were looking for feedback on tags that he thought were missing or present but unnecessary, as well as those which he felt he would not have the time or knowledge to use.

Through our collaboration with Fontanella and the NAL staff, we were able to develop a final tag list to use in the transcription of the Robinson Reports. This list was comprised entirely of tags from the TEI and CIMI standards, and did not include any custom tags.

3.5.3 Defining the Document Type Definition (DTD)

The final step we took in the development of tags for transcribing the Robinson Reports was defining the DTD that Professor Fontanella would use to tag the reports. This DTD would contain all of the TEI tag sets and CIMI tags that we would be using. It would also define the rules that delineate the structure of Professor Fontanella's transcriptions of the Robinson Reports.

The first step we took in defining the DTD was compiling the core, base, and additional TEI tag sets that we would be using. From our previous research and comparisons between our tags and the standard lists, we had already developed an idea of which tag sets would need to be incorporated into the DTD. We continued our research, looking at the content and purpose of various tag sets that appeared to be pertinent. Of these, we decided which would be necessary to include in our final tag list.

Once we decided which tag sets to include, we had to merge them. We accomplished this by using Pizza Chef, which is an internet site specifically designed to merge TEI tag sets. Pizza Chef merged together all of the elements in the various tag sets, ensuring that they followed proper TEI structure. There were a few problems present in the DTD that Pizza Chef formed, which we discovered and were able to fix. With that done, we had a well-formed DTD containing all of the TEI tag sets that are necessary for the transcription of the Robinson Reports.

At that point, we had to introduce the required CIMI tags to the DTD. This was a very tedious process that required us to go through all of the existing elements in the DTD and add the new tags. The DTD determines which tags can be inserted within other tags. For instance, the DTD allows bibliographic information to be inserted in the header, but not into the body of the text. We had to determine the function of each existing tag in the DTD and decide which, if any, tags should be allowed to be inserted inside it.

Once all of the necessary TEI and CIMI tags were included in the DTD, we compiled it and began testing it. Thorough testing was necessary to ensure that we had properly defined all of the tags. Using the XML editor software that we chose for

Fontanella's transcriptions, we created documents that used the newly defined tags in various ways. This would ensure that they were functioning properly and obeying the structure that we desired.

3.6 Creation of Transcription Templates

Once we had developed a working DTD containing all of the necessary tags for marking up the Robinson Reports, we were able to start developing the templates for Professor Fontanella's transcriptions. These templates will provide the basic structure for the transcriptions and a custom environment for editing and viewing the documents.

The templates will determine the basic structure for the transcriptions by requiring that the document meet the DTD specification as well as providing the necessary tags for the underlying report that it represents. As part of this structure, the complex TEI header, required by the TEI specifications, will be predefined and automatically inserted at the beginning of each transcription. When Fontanella creates a new document he will be prompted for the information required to define the header.

The templates will also provide a custom environment for editing and viewing the documents by specifying fonts, menus, toolbars, page layouts, and special formatting. The custom environment settings only pertain to viewing the document in the editing software and do not define how other XML viewers display the document. This view of the document and all of the special formatting is meant to ease Professor Fontanella's transcriptions by providing a consistent structure for transcribing the reports.

3.6.1 Determining Document Types

In the process of creating transcription templates, we needed to first determine which document types required templates and how much detail the templates needed to include about the original document type. One of our main concerns was whether or not to create a different template for various versions of similar reports.

We discussed the various forms of each document type with John Meriton. In this discussion we decided which differences were important to note in the transcriptions. In order to analyse the differences, we needed to compare numerous versions of the same document type. This required us to construct a list of similarities we found between the different versions of the same document type and to determine if the differences were significant enough to take note of and include in a template. For instance, two versions of minute paper only have one difference in their printed headings. One minute paper has "South Kensington Museum" in it heading and the other has "Science and Art Department." Apart from these headings, the two types of minute paper are exactly the same. Mr. Meriton felt that this difference was important to differentiate because the change in the headings signified a change in the museum's governance.

3.6.2 Developing the Templates

After deciding on a final list of document types, we researched the software that we had chosen and determined how to create SGML/XML templates. This involved reading the manuals, taking built-in tutorials, and examining sample templates included with the software. Once we were familiar with the software we began designing templates for the various types of Robinson Reports.

The first task that we needed to complete before constructing the templates was to meet with Christopher Marsden, the assistant archivist for the V&A, to discuss the appropriate bibliographical information that must be included in each transcription. This meeting led to the development of a system to provide a bibliographical reference for the transcriptions. At this meeting we also learned the appropriate bibliographic citation for a particular Robinson report. This bibliographical information will be included in the header of each transcription.

Our next step was to go back to the examples that we had recorded in our analysis of the Robinson Reports and examine the format of each type. Based on the examples, we developed a generic template that represented the reports of that type. We decided on the fields that should be represented in each template and determined how to tag them appropriately. This required careful design and constant testing. Since formatting is not provided by XML, we needed to develop a way to overcome the problem of associating information with other information that was related in the original report through positioning on the page. For example, one of the document types that we found, the professional reference paper, has two columns: one for the description of the item and the other for the item's date and country of origin. In order to link these two pieces of information we needed to do more research into the capabilities of the tags being used and into the restrictions imposed by XML.

3.7 Sample Transcriptions of the Robinson Reports

Upon completion of the document templates, we entered a sample transcription into each of the templates we developed. These transcriptions provided additional testing of the templates. Moreover, they supplied us with an estimate of how long it will take

Professor Fontanella to transcribe the average letter and served as examples for Professor Fontanella. Ideally, it would have been more beneficial to have Fontanella do the sample transcriptions himself, but this was infeasible because he could not come to London for testing.

When we began our transcriptions, we chose a report that was representative of each template and attempted to transcribe it in a manner that we felt would be similar to Fontanella's work. We tried to choose reports that would use as many different tags as possible, which increased the likelihood that we would find a problem. Then, while entering the transcriptions, we took note of any bugs or inconsistencies that we discovered. We also noted any stages of the transcription process that were difficult to understand or carry out. In addition, we recorded the time that it took us to complete these sample transcriptions to assist in our estimation of how many of the Robinson Reports Fontanella will have time to transcribe. We did not, however, attempt to fully test the templates for usability, as we were too familiar with the tags and templates. This would be better accomplished by people with less background in the project.

The completed sample transcriptions are intended to aid Fontanella when he is learning how to use the software and templates. They will provide a reference for him to determine which types of reports are associated with which templates. They will also help him visualise where the various parts of the transcriptions fit into the templates.

After we completed the sample transcriptions, we revised the tags and templates to reflect the problems that we found. We determined which of these problems were significant and could be fixed, as well as which were trivial or could not be overcome. Some of the problems were impossible to fix due to the limitations of XML or of the

editor. Of the significant problems, we addressed those that could be fixed, and documented those that could not.

3.8 Development of a User Manual for Transcribing the Reports

After testing the templates and tags by conducting sample transcriptions, we wrote a user manual that describes each template and its use, provides a detailed explanation of each tag, and lists known problems and bugs. This document also contains instructions on how to use the SGML/XML editor that we chose. Although we will explain all of this to Professor Fontanella in person before he arrives in London, this document will serve as a reference while he is actually doing the transcriptions.

The description of each of the templates includes a reference to an actual report that is associated with that template. Professor Fontanella will be able to use these example reports to determine which reports are related to each template. In addition, the user manual was designed to eliminate confusion about the significance of each tag. The descriptions of the tags that Fontanella will be using will help him determine where to insert them, as well as what the various attributes mean. These sections are supplemented with the list of known bugs and problems in the templates and tags. This list was drawn from our sample transcriptions, and reflects problems that were inherent in the software and in XML that could not be fixed.

Also included in the user manual are instructions on how to use the SGML/XML editor to provide Fontanella with a reference on how to load a template, insert tags and entities, enter text into the templates, save his work, and other simple tasks. This part of the user manual does not include extensive documentation on how the editor works, because that information is available in the online help and tutorials. Rather, it contains

information specific to the tasks that Fontanella will need to do while transcribing the Robinson Reports.

This user manual will not only be of benefit to Fontanella, but will also serve as a reference for the team designing and implementing the electronic search system. They will be able to use the document as a reference to the various tags and document types that the system will need to recognise.

3.9 Final Testing

After the completion of the user manual we tested its effectiveness and gathered outside feedback on the usability of the templates and tags. This allowed us to further refine the templates, tags, and user manual because it brought a point of view from people with no prior experience with the templates, tags, and editor.

Our IQP advisors, Professors Chrysanthe Demetry and Richard Vaz of WPI, volunteered to transcribe some of the Robinson Reports using our templates and tag list. Demetry and Vaz were good candidates because they do not have a significant background in SGML or XML. They were given the exact resources that Professor Fontanella will be working with to transcribe the reports, including the user manual and the editor set up with the templates and completed DTD. We then chose letters for them to transcribe, making sure that they were legible and contained sufficient information for them to tag.

Once completed, Demetry's and Vaz's transcriptions and observations were very useful because they allowed us to see how intuitive our templates and list of tags were. They were asked to give us feedback regarding the layout of the templates, any difficulty in understanding or using the tags, and any bugs or mistakes that they discovered. We

also examined their finished transcriptions to see if they had been correctly entered and fully tagged all the information. This helped us to locate any areas that we did not emphasise enough or clearly explain. Once we thoroughly checked Demetry's and Vaz's transcriptions, they were added to ours' to serve as further examples for Fontanella.

3.10 Estimation of Transcription Time

The NAL expressed a desire for us to estimate how many of the Robinson Reports Professor Fontanella will have time to transcribe. In addition, we decided to estimate how quickly Fontanella would need to transcribe each page in order to finish the entire collection.

To do this, we needed to estimate the number of pages in the collection. We made our estimate based on the number of pages and not the number of reports, because the number of pages per report varies so greatly that it was not possible to get a good estimate of the number of reports. Next, we contacted Fontanella and asked him how long he would be staying in London and how much time he planned on spending working at the NAL. We also asked how much time he was planning on spending adding additional comments on the content of the letters. Using this information and the average time that it took Professor Vaz, Professor Demetry, and us to do our sample transcriptions, we were able to estimate both the portion of the reports that Fontanella would have time to transcribe and the average amount of time he would need to complete each page if he were to transcribe the entire collection.

To find out approximately how far Fontanella would get, we calculated the total number of minutes that he would be working at the NAL based on the information that he provided. We took the total number of minutes and divided it by the average time per

page that it took us to do our sample transcriptions. This gave us the estimated total number of pages that Fontanella will be able to complete. Dividing this number by the estimated number of pages in the collection gave us the approximate percentage of the Robinson Reports that Fontanella would finish. To estimate how quickly Fontanella would need to transcribe each page in order to complete the entire collection, we simply divided the amount of time he would be working by the estimated number of pages.

3.11 Development of Guidelines for Completing the Search System

To further aid the team that will design and implement the search system, we prepared a detailed description of the system they are meant to develop. In this document, we make recommendations on what features and capabilities should be incorporated into the system. These include the different ways in which users can search the letters, the information that should be included in the list of search results, and the information that should be included in the display of the letters. Some recommendations have been made regarding the manner of display as well; however, this will be left largely up to the system designers.

Our recommendations were drawn from all aspects of our work with the Robinson Reports. We used our own experience, drawing ideas from our analysis of the reports and our knowledge of TEI and XML. More importantly, however, we designed our recommendations around our conversations with Professor Fontanella and the NAL staff. Throughout the course of our work, we had many discussions with them regarding the types of information the system should search for, how the reports should be displayed, and how they should be catalogued.

By following these guidelines, a follow-up team should be able to create a system that scholars can search for specific information, which has been chosen and evaluated with the help of scholars. This system should be able to interpret the modified TEI format in which the Robinson Reports will be held and properly display them to the user's screen. However, because of the use of standard non-proprietary formats, the reports will not be locked into this system; they will be portable to any other system or application that supports their format.

4 Results and Analysis

This chapter presents the results of our work with the Robinson Reports. It includes the findings from our analysis of the Robinson Reports, the editor we chose, the final tag set, a list of the shortcuts to common phrases, a description of the document templates and sample transcriptions, as well as the estimate of the amount of time it will take to do the transcriptions.

4.1 Analysis of the Robinson Reports

During our initial analysis of the Robinson Reports, we developed lists of the unique formats and styles of documents, important information that may be tagged, and the common phrases that we found. These lists represent the information that we drew from the Robinson Reports, and have not been streamlined to reflect our conversations with Professor Fontanella and the staff of the NAL. They do, however, give explanations that describe the decisions we made regarding each.

4.1.1 Unique Formats and Styles

The unique formats and styles that we decided upon while reading the reports represent the various types of documents within the collection. A complete list of the formats and styles we identified is available in Table 2. This list includes an explanation of each type, as well as an example. The examples are given in the form (binder number)-(letter number). The binder numbers take the form of MA/3/#, where MA stands for Museum Archive, 3 represents the Robinson Reports, and # represents the specific binder number. The letter number represents the number of the sleeve where the

first page of the report is held. These numbers run consecutively throughout each volume, where a volume represents approximately four binders.

If more than one example is listed, the document has varying forms. For instance, minute paper has several different examples listed. This is because it is a pre-printed form that, over the years, changed slightly in its design. In one of the examples, the header on the minute paper includes "Science and Art Department." Several years later, this part of the header was changed to "South Kensington Museum." This change is important to note in the transcriptions of the Robinson Reports. Some forms had much less significant changes, such as a change in stationery numbers, which will not be noted in the transcriptions.

Table 2 – Document Types from the Analysis of the Robinson Reports

Document Type	**Explanation**	**Example(s)**
Letter	The letter type represents a standard letter, containing a date, a body, and a signature. Letters come in various sizes and lengths, but all follow the same basic format. Occasionally, however, they include variations. These could be in the form of a greeting, a numbered or lettered list within the body of the letter, or a summary of the letter's contents. These summaries were written on the letters in red by a third-party.	MA/3/1-27
List	Lists represent reports that are pure lists, not lists included within the text of a letter. Examples of lists in the Robinson Reports include objects in the museum, objects purchased for the art library, and objects with registry numbers and prices.	MA/3/11-349 MA/3/11-345
Minute paper	Minute paper is a pre-printed form, which was used to acknowledge and give directives regarding other reports. This form includes three columns, one for names, one for dates, and one for minutes. However, these columns were often ignored, and had all three parts written in the minutes column. Minute paper has several variations.	MA/3/12-16 MA/3/46-71

Table 2, cont'd – Document Types from the Analysis of the Robinson Reports		
Document Type	**Explanation**	**Example(s)**
Telegraph	Telegraphs represent the various types of telegraphs that were used. Although there are a few different types, sent from different companies, all the telegraphs contain the same basic content.	MA/3/16-47 MA/3/16-50 MA/3/23-216
Printed materials and sketches	Printed materials and sketches include any materials included with the reports that are not some form of hand written correspondence, either to or within the museum. These include printed catalogues of sales, full-page sketches, postcards, and printed pictures. These items either cannot or do not need to be transcribed.	MA/3/10-280
Professional Reference Paper	Professional reference paper is a form letter that is very common beyond MA/3/28. These forms are specifically designed for reporting on works of art. They have separate fields for the name of the object, the asking price, the object's date and country of origin, and the body of the report. The body includes the Art Referee's recommendation regarding the object's purchase. Later variations of this form also include a field for the seller's name.	MA/3/28-71 MA/3/46-127
Preliminary reports for the information of the director	Much like Professional Reference Paper, these reports are structured forms designed specifically for reporting on art objects. This form was used only for a short time, appearing in MA/3/28.	MA/3/27-47
Form letter	The form letter reports are pre-printed letters with a few blank spaces left for the recipient's name, the topic of interest, and a signature. These reports referred to particular objects that were being considered for purchase. These objects were listed below the signature of the writer.	MA/3/46-97/1
Miscellaneous	Miscellaneous reports include those that are not covered by the other types. All of the Robinson Reports should be covered in some manner by one of the types previously listed; however, a completely free-form document is necessary for any that are not.	None

There are two types of information that appear on most or all of these types of documents. The most important of these are minutes and annotations. Minutes and

annotations are comments, in various forms, that were written on the reports by their various readers. These minutes range from long replies, giving directives on how the report should be handled, to simple initials, signifying that it has been read. There are often many minutes on a report, sometimes occupying more space than the report itself. On occasion, the minutes even spilled onto the back of the paper or onto an additional sheet.

The second type of information that is common to many of the document types is an introductory summary of the report. Once the report was folded, and presumably sealed, a brief summary of its contents was sometimes written on the outside. These summaries include a date and signature, causing them to resemble a small letter. Because the reports are now unfolded, the summaries appear sideways on the backsides of the reports that have them. The reports with these summaries include letter, minute paper, and professional reference paper; they are not, however, common among these types.

4.1.2 Information to be Tagged

The list of information that we identified for tagging is made up of information that we thought would be important to scholars, and could be recognised by the final search system. A complete list of the types of information that we found and the corresponding tags we identified is available in Table 3. This list is a complete list of information that we drew from our analysis of the Robinson Reports, although tags that are not being used are marked in italics. This table also gives an explanation of each type of information, including any corresponding TEI or CIMI tags that could be found.

Table 3 – Key Information from the Analysis of the Robinson Reports	
Type of Information	**Comments**
Binder number	This number identifies which binder the report being transcribed is located in. The binder number will be included in the bibliographic information in the TEI header. It will be marked with the <biblScope> tag, with the "type" attribute set to "reference."
Letter number	This number identifies the number of the letter within each volume of reports. The letter number will be included in the bibliographic information in the TEI header. It will be marked with the <biblScope> tag, with the "type" attribute set to "item."
Registered paper number	This is the number stamped on the report and recorded in the register. Not all of the reports have this number. When present, it will be included in the bibliographic information in the TEI header. It will be marked with the <biblScope> tag, with the "type" attribute set to "superceded"
Name	There are several types of names that are of conceivable importance. Among these are: artists, sellers, previous owners, and reproducers of art; senders and recipients of the reports; readers who did and did not leave comments; and recipients of comments. All of these names can be marked with the <persName> tag. The specific types of names that we chose, along with any others that become necessary, can be specified in the "type" attribute.
Date	There are several types of dates that may be of importance: the date the letter was written, received (i.e. stamped), or commented on; the date an artefact was purchased; and the date of an artefact's origin. The tag <date> exists, but has no attributes; therefore it cannot specify different types. The date a report was written can, however, be signified by enclosing it in the <opener> tag.
Geographic location	Many types of geographic locations can be found in the Robinson Reports, including cities, countries, and regions. These locations are all covered by the <placeName> tag, and can be more specifically named by its subelements. Specific places, such as museum and auction house names also come up. The <orgName> tag can be used for these places.

Table 3, cont'd – Key Information from the Analysis of the Robinson Reports	
Type of Information	**Comments**
Art object name and descriptive information	The name of an artefact will be designated with the CIMI tag <objectName>; the type of object will be set in the "type" attribute of this tag. The places of the object's origin and purchase will be tagged as a <placeName>. The CIMI tag <materialName> will denote the material that the object is made of. The price of the object will be tagged with the CIMI tag <cost>.
Language of text	There is a "lang" attribute for each tag, which is used for marking a whole section or element as foreign. The <foreign> tag is used for a phrase or word in another language.
Museum object number	The CIMI tag <objectIdentifier> will be used to mark museum object numbers.
Art collection	This would signify the name of a specific collection, such as the Robinson Reports. The CIMI tag <collectionName> will be used to mark names of collections.
Price of collection	*The price of a collection was found to be too insignificant to be tagged.*
Lot number in an auction	*This would denote the lot number of an item in the auction it was being sold in. This information was decided to be extraneous and will not be marked.*
Size of Art Object	*It may have been significant to mark the size of an art object by base, width, height, and diameter. This information was decided to be irrelevant and will not be tagged.*
Pre-printed text	This would be used to signify printed text, such as the headings on minute paper. The <hi> tag, with the "type" attribute set to "printed," will be used to signify pre-printed text.
Pasted on pre-printed material	This will be tagged the same way as other pre-printed text, using the <hi> tag with the "type" attribute set to "printed."
Stamp	*This would signify the Department of Art and Science stamp, which holds the registered paper number. The registered paper number will be recorded as noted above, but the stamp itself will not be tagged.*
Copied letter	*This would signify text that was a copy of the original (these are hand-written copies, not photocopies). This information will not be tagged.*

Table 3, cont'd – Key Information from the Analysis of the Robinson Reports	
Type of Information	**Comments**
Picture enclosed with a letter	*This could be a printed picture or a sketch drawn by the author of the report. Pictures can be inserted via the <figure> tag, with a caption provided by <figDesc>. However, this would require the digitising of the images. Although this may be done in the future, the images will only be described in <note> tags for now.*
Item on loan	*This would signify the name of an item on loan to the museum. As with other artefacts, these will be signified with the CIMI tag <objectName>; however, they will not be specified as being on loan.*
Copied artefacts	*This would include casts, photographs, and other reproductions. As with the items on loan, these artefacts will be signified as such with the <objectName> tag, but will not be specified as copied.*
Text signified by a reader	This could included text underlined, pointed to, or otherwise signified. Underlined text can be imitated with the <hi> tag; text that is pointed to can have any associated notes linked to it, but the pointers themselves will not be imitated.
Registered paper number in the text	This would be used when the text of a report mentions the number of another registered paper. These registered paper numbers will not be marked in the text, but the referenced reports will be noted as related letters.
Signature	The signature of the report's author will be tagged as <signed> with the <closer> of the text.
Registered number of document	*This is a field on Professional Reference Paper. It is rarely filled in and we could not determine the significance of the number. This information was found to be irrelevant and will not be tagged.*
Private	*Some letters have parts signified as private. This information will not be tagged.*
Crossed out or added text (i.e. a fixed mistake or added word)	The tags <add>, <del>, and <restore> are all used to describe different types of internal editorial changes.

Table 3, cont'd – Key Information from the Analysis of the Robinson Reports	
Type of Information	**Comments**
Text written in by other than author	This text is generally in the form of minutes and annotations written in response to the report. Less often, it is text that was added to the content of the report. In the former case, the <note> tag will be used; in the latter case, the internal editorial tags listed above will be used.
Paper size	*The paper size was found to be irrelevant and will not be tagged.*
Paper colour	*The paper colour was found to be irrelevant and will not be tagged.*
Text colour	A change in the colour of the text (i.e. a change in the ink) can be signified by the "ink" attribute in the <handShift> tag.
Illegible	Material can be illegible for many reasons; poor handwriting, smudges, tears in the paper, burn marks, and stains are just some forms of illegibility. The tags <unclear>, <damage>, <gap>, and <supplied> provide for marking varying degrees of illegibility.
Link to other letters	Reports are referenced in and related to each other for a variety of reasons. As the process of linking is quite complex, these letters will be recorded and linked at a later date. The tags <link> and <xptr> will most likely be used for this linkage.
Footnote made by author	Some of the reports have footnotes in them, which would need to be included, but still be designated as footnotes. The <note> tag would allow for this.
Footnote made by transcriber	This allows the transcriber to include links to comments on the text. These would also use the <note> tag, being differentiated via the "type" and "resp[onsibility]" attributes.

It is important to note that some of the types of information we identified do not have corresponding standard tags; however, we found no information that did not fit into TEI or CIMI that was important enough to deviate from the standard. This was decided through discussions with John Meriton, the head of public relations at the NAL, and Douglas Dodds, the computer specialist.

In addition, it is important to mention that this is not a complete list of tags that will be used to transcribe the Robinson Reports. Many more tags from the standard will be used to determine the structure and format of the reports. Some of the standard tags will also allow Fontanella to add his own editorial notes to the transcriptions. For a complete list of the tags that will be used in the transcriptions of the Robinson Reports, see Section 4.3.

4.1.3 Common Phrases

In order to facilitate the transcription process for Professor Fontanella, we felt that it would be beneficial to have shortcuts to phrases that appeared commonly throughout the reports. These phrases ranged from signatures made by management to the most commonly used closing of a letter. These common phrases can be seen in Table 4.

Table 4 – Common Phrases from the Analysis of the Robinson Reports

Common Phrase	**Explanation**
J C Robinson	At the end of almost all of Robinson's reports he would sign the letters J C Robinson
Mr. Cole	*This was the most common way Henry Cole would sign reports. In almost all of the reports, Henry Cole would sign the letter signifying he had read it and sometimes he would supply a comment explaining whether he wanted to purchase the object or not.*
Mr. Digby Wyatt	This is the name of one of Art Referees who worked for the museum upon Robinson's departure.
Rick Redgrave	*This man was another man who commonly signed Art Referee's reports. He would give his specific comment about an item or items discussed in the report.*
Mr. Whitehead	*This man was discussed throughout several of the letters.*
South Kensington Museum	Throughout several of the reports the South Kensington Museum is discussed. One of the most common places it is discussed is when an Art Referee suggests whether an object should be bought.
Kensington Museum	This phrase is also used commonly by Art Referees when discussing an object's purchase.
Art Museum	*Throughout the reports Art Referees referred to the museum in several ways, but the generic phrase "Art Museum" is a part of many of them.*

Table 5 – Common Phrases from the Analysis of the Robinson Reports	
Common Phrase	**Explanation**
Art Library	*Sometimes in the reports one of the Art Referees will recommend an item for the Art Library specifically.*
Science and Art Department	This was the name of the governmental department that ran the museum at the time.
Your Obedient Servant	This was one of the most common closings used in reports.
Your Most Obedient Servant	This is another popular closing for a report.
See minute	*This phrase was used when referring a letter to its specific minute paper.*
See letter book	*This phrase was used to refer the reader to the letter book for more information.*

These common phrases were eventually incorporated into the document templates via QuickWords. Quickwords allow the user to type an abbreviated form of the phrase, which is automatically expanded. Some of the common phrases were decided to be too short or insignificant to make QuickWords worthwhile. These phrases are signified in italics in Table 5. All of the other phrases had QuickWords defined for them.

4.2 Selection of an Editing Environment

We chose WordPerfect 9 as the editor for the transcription of the Robinson Reports. Our research into various SGML/XML editors led us to WordPerfect 9 because it would be the most comfortable editor for Professor Fontanella to use. Fontanella has used a previous version of WordPerfect as a word processor; thus, he is already familiar with the program's general layout and use. Moreover, WordPerfect 9's capabilities were best suited to our needs

WordPerfect 9 has the capability of creating and editing SGML/XML documents. This capability allows for the enforcement of our tag set and its rules while Professor Fontanella is transcribing the Robinson Reports. The editor will prevent tags from being

inserted where they do not belong and will require tags to be inserted where they are required. For example, it will only allow a salutation tag to be put in the opener or closer of the body.

In addition, WordPerfect 9 allows for the creation of templates, which will enable us to lay out a set of tags that must be filled in for each document type. This will provide Fontanella with a consistent format for transcribing the Robinson Reports. WordPerfect 9 templates also allow for the creation of hotkeys, which insert predefined text when they are executed. These hotkeys can be used to insert the common phrases we found in our analysis of the Robinson Reports.

4.3 Final Tag Set

While conducting our analysis of the Robinson Reports, we were researching and developing a final tag set. This final tag set took the form of a document type definition (DTD), which defines the tags and the rules that will specify the structure of the transcriptions. This section contains a list of all the tags in the final DTD constructed for the transcription of the Robinson Reports.

The DTD was compiled from the TEI core tag set, prose base tag set, five additional tag sets, and six individual tags from the CIMI standard. The core tag set is mandatory for any TEI compliant text and provides general tags that can be used by all types of documents. The core tag set also contains tags used in the TEI header, which holds information about both the electronic and source versions of the text. The prose base tag set implements a part of the core tag set that defines the structure of the actual text. The additional tag sets contain tags used for more specific applications that would

not be necessary in all types of documents. The tags imported from the CIMI standard deal with information specific to museums that is unavailable in TEI.

The additional TEI tag sets that were included were *Names and Dates*; *Tables, Formulae, and Graphics*; *Linking, Segmentation, and Alignment*; *Simple Analytic Mechanisms*; and *Transcription of Primary Sources*. The purposes of the first two of these are fairly self-explanatory. *Linking, Segmentation, and Alignment* provides tags and methods for creating references within a single text and between multiple texts. *Simple Analytic Mechanisms* allows text to be broken up into its grammatical components for analysis. *Transcription of Primary Sources* provides transcribers of hand-written material with tags for editorial changes and other markings.

Two TEI extension files, *VAExt.ent* and *VAExt.dtd*, were developed to incorporate the CIMI tags into the TEI tag set. An extension file is a document that contains any additional tags, additional entities, or modifications to existing tags in the TEI tag set. These two files were developed following the guidelines provided by the TEI documentation. They define the CIMI tags that were added to the DTD, as well as modifications that were made to the existing TEI tags to incorporate the CIMI tags. These two files can be found in Appendix D.

For presentation here, the tags in the DTD were separated into those that Fontanella will be using to mark up the text of the Robinson Reports; those that will be used in the transcriptions, but have already been inserted via document templates; those that may be used to mark the documents more explicitly in the future; and those that are irrelevant to the Robinson Reports. This last set of tags will not be explained here, but is

available in Appendix D. All of the definitions given in this section were adapted from the standard definitions given in the TEI and CIMI documentation.

4.3.1 Tags Necessary for Marking Up the Text

The tags discussed in this section are those that Fontanella will need to insert manually to mark up the text based on his findings. These tags determine the structure and style of the text. The latter of these could entail the physical appearance or the intended meaning of the text. Finally, this section includes tags used for marking content-specific information, such as names and dates. All of the tags presented in this section are described in more detail, and with full examples, in the *User Manual for Transcribing the Robinson Reports* included as Appendix F.

4.3.1.1 Structure and Grammar

The tags in this section are used specifically for noting the structure and grammatical marks used in the text. It is important to note that, with the exception of quotation marks, punctuation is handled internally by the editor and can be entered as if the transcriptions were being done in plain text.

- **<p>** signifies a paragraph. <p> tags are generally used between the <opener> and <closer> of the <body> in a document. Most of the other elements in this guide will be used within a <p> tag. <p> tags are not generally nested within one another, but many can be used sequentially.

- **<note>** could contain a minute or annotation on the original document or a note from an outside source, such as the transcriber.

- **<quote>** marks a quotation attributed to an outside source. <quote> should be used in place of quotation marks in the transcriptions.

- **<pb>** and **<lb>** signify a page break and a line break, respectively.

4.3.1.2 Lists

These tags are specifically associated with creating lists. A list in TEI can be numbered, lettered, or ordered in any other manner. Moreover, lists can have two columns: the main list item column and a label column. Both columns can, but are not forced to, have headings. This can be done with the <head> tag described above, or the <headLabel> and <headItem> tags described below.

- **<list>** is the main element of this type. It is used to denote a list of any type, and contains the entire list.
- **<item>** is a sub-element of <list> that contains a single item in the list.
- **<label>** is a sub-element of <list> that signifies the label for a list item. For example, the term being defined in a glossary would be tagged as a <label>.

4.3.1.3 Physical Appearance of Text

Whereas the last set of tags was used to denote the structure of the document, these tags are used to describe its physical appearance. They allow the transcriber to specify a change in the style of the writing or a section of text that is illegible. Moreover, the transcriber can demonstrate the cause and degree of illegibility by choosing different tags.

- **<handShift>** is used to mark a change in hand, writing style, or ink.
- **<unclear>** is used to indicate text that is not fully legible and cannot be transcribed with absolute confidence.
- **<damage>** signifies text that cannot be transcribed with certainty due to damage of any kind.
- **<gap>** indicates where material has been omitted in a transcription, for any reason. <gap> should be used if the transcriber cannot make out a given portion of text well enough to make a guess at what it says.

- **<supplied>** can be used in place of <gap> if the transcriber enters invented text in place of original text that is completely illegible. This text represents the transcriber's best guess as to what the text originally said.

- **<space>** is used to designate a noticeably large amount of space in a line of text. The <space> tag allows the transcriber to provide the direction and extent of the space.

4.3.1.4 Internal Editorial Changes

The tags in this section are used to denote editorial changes in the original document. These changes could include text that was crossed out by the writer, or written in between the lines. It could also include editorial changes by later readers of the documents, if these changes are indicated on the actual document. These tags are not properly used for editorial changes by the transcriber. For text that was deleted thoroughly enough that it cannot still be read, see the <gap> tag in the previous section.

- **<del>** indicates text that was deleted in the original document. This could include text was crossed out, or otherwise removed, but can still be read.

- **<add>** indicates text was added in the original document. This could include words written between the lines or in the margins.

- **<restore>** signifies text that was restored to an earlier state by the author or a later editor. For example, text that was crossed out and then marked to show that it should be kept would be marked with <restore>.

4.3.1.5 External Editorial Changes

The tags included here are used for marking abbreviations, for which the transcriber can note the expanded form; incorrect text in the original document, for which the transcriber can include both the correct and incorrect form; and irregular text, for which the transcriber can include a regularised version.

- **<abbr>** could contain any form of an abbreviation.

- **<sic>** signifies text that is known to contain a mistake.

- **<orig>** is used to contain text that is antiquated or otherwise abnormal. This could include outdated spellings and usage of words.

4.3.1.6 Text Style and Tone

The next segment of the tag list is comprised of tags that focus on the style and tone of the text. These tags are used to mark text that has special meaning that may not be apparent, or that may require added interpretation.

- **<hi>** signifies text graphically distinct from surrounding text. This could be text in a different font or text that is printed as opposed to written.

- **<emph>** signifies text that is stressed or emphasised in a linguistic, not typographic, sense.

- **<distinct>** is used to mark linguistically distinct words or phrases, such as slang, technical or archaic terms.

- **<foreign>** designates text as being in a different language from the rest of the document.

- **<soCalled>** is used to designate words or phrases that are not normally used and the author does not take credit for. For instance, a word in "scare quotes" would be tagged as <soCalled>.

- **<mentioned>** marks a word that is mentioned in the text, but is not used.

4.3.1.7 Names, Dates, and Numbers

The final set of tags that Professor Fontanella will be using to mark up the bodies of the Robinson Reports are references to specific aspects of their content. Included here are tags used for marking dates, numbers, measures, and various types of names.

- **<persName>** is used to mark a person's entire name, including titles, forenames, surnames, and generational information.

- **<placeName>** indicates the name of a place, whether it be a geographic formation or a name of specific geo-political region. Also, if the writer is referring to an area near a specific landmark, any distance or directional

information relating the area of interest to the landmark would be included in the <placeName> tag.

- **<orgName>** is used to contain the name of an organisation of any sort. This could include private corporations, public organisations, and government agencies.

- **<objectName>** denotes the name of an object, such as a piece of art or an item in a museum.

- **<collectionName>** is used to mark the name of a collection, as opposed to individual pieces.

- **<materialName>** signifies the material that an object is made out of.

- **<eventName>** specifies the name of a specific event, such as the World's Fair.

- **<name>** indicates a proper noun. <name> will only be used to indicate names not covered by the more specific tags listed above. In particular, <name> will be used with the "type" attribute set to "site" to tag a specific place, such as the name of a telegraph station.

- **<date>** is used to mark any manner of date.

- **<dateRange>** is used to tag a time range. This could be a specific mention of two dates or a general time range, such as "15th Century."

- **<objectIdentifier>** denotes an identifying number for an object; in the V&A this would be the museum object number.

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

- **<num>** is used to mark any form of a number. For example, this could mark a number written out, in digits, or in roman numerals.

- **<measure>** refers to any quantity. This tag usually contains a number, a unit, and the name of item being measured.

4.3.2 Tags Used Only in Document Templates

The following tags appear only in the document templates, and will not need to be inserted at any time by Professor Fontanella. For more complete information on how these tags apply to the templates and how data should be entered into each within the

various templates, see the *User Manual for Transcribing the Robinson Reports* included as Appendix F.

- **<TEI.2>** contains the entire TEI document. This tag signifies that the file is in the TEI format.

- **<teiHeader>** contains a description of the text held in the file. The header information is affixed to the beginning of every TEI text, and contains specially formatted information about the original and electronic documents, the encoding of the document, and its revision history.

- **<teiCorpus>** is used to group a collection of texts together in one large file consisting of a collection header and the body of the individual texts. The individual texts still retain their own headers.

- **<text>** contains an entire text, including the actual body and any front and back matter it may have. Front matter includes prefaces, tables of contents, and other materials that appear previous to the text of the document. Back matter includes indices, appendices, and other material that appears after the text of the document. <text> can be used to mark either a single text or a composite of texts.

4.3.2.1 The TEI Header

There are four main tags that can appear within the <teiheader> in a document: <fileDesc>, <encodingDesc>, <profileDesc>, and <revisionDesc>. Of these, <fileDesc> is the only mandatory element; however, <profileDesc> is used in the Robinson Reports header as well.

- **<fileDesc>** contains bibliographic information regarding both the electronic text and the source material.

- **<profileDesc>** contains classification information, as well as a description of the text's subject matter.

4.3.2.1.1 The File Description

All of the tags in this section are used within the <fileDesc> element in the TEI header. These tags are used to provide bibliographic information about the electronic version of a text, as well as about the source.

- **<titleStmt>** is a mandatory sub-element of <fileDesc> that contains the title of a work, as well as information describing the people who are intellectually responsible for it.

- **<title>** is a sub-element of <titleStmt> that denotes the title of a work, including any subtitles. This can also be used as a sub-element of <biblStruct>, described later in this section.

- **<respStmt>** is a sub-element of <titleStmt> that specifies a person intellectually responsible for work. It is generally broken up into two parts: the <resp> tag, described below, and a <persName> tag that contains the responsible person's name. <respStmt> is used for those that do not fit into the predefined categories of <author>, <editor>, <sponsor>, <funder>, or <principal> (as they are irrelevant to this project, these tags are not documented here). For instance, a transcriber would be indicated in the <respStmt> tag.

- **<resp>** is a sub-element of <respStmt> that contains the manner of intellectual responsibility that the person holds. For instance, the phrase "transcribed by" would be placed in the <resp> tag.

- **<publicationStmt>** is another mandatory sub-element of <fileDesc>. <publicationStmt> contains information regarding the publication and distribution of the electronic text.

- **<authority>** is a sub-element of <publicationStmt> that indicates the name of a person or organisation that provides the electronic text.

- **<sourceDesc>** is the final mandatory sub-element of <fileDesc>. It contains bibliographic information regarding the source material that the electronic text was based on.

- **<bibl>** denotes loosely structured bibliographic information, such as a textual citation. This can contain the <title> and <biblScope> elements, as well as various tags for names and publication information.

- **<biblScope>** is used to show the scope of the item being cited. This could include a volume number, part number, and item number.

4.3.2.1.2 The Profile Description

All of the tags in this section are used within the <profileDesc> element in the TEI header. These tags are used to give the document a type, such as "minute paper" or "letter." The electronic search system will use this information to choose which display method it should use.

- **<textClass>** is used to group together all information regarding the classification scheme that will be implemented. This includes information about the type of document.

- **<keywords>** a sub-element of <textClass>, this tag designates a list of keywords that describe the document or explain its format.

- **<term>** denotes a technical term. This tag is used to signify the document type within the <keywords> tag.

4.3.2.2 The Text

The following tags are used within the <text> of the document templates, but will not have to be used at any other time in the transcriptions of the Robinson Reports.

- **<body>** contains body of a text; this pertains to only a single text, not a composite. Also, <body> does not include any front or back matter.

- ** ** contains any form of subdivision of the front, body, or back of a text. This could include chapters, sections, or any other division larger than a paragraph. Here is used to delineate columns of text.

- **<head>** contains a heading of any type. For instance, it can be used as the header for the body of a text, or for a list or table within the text.

- **<opener>** contains any opening information, such as a date, salutation, or byline.

- **<closer>** contains any closing information, such as a salutation, signature, or dateline.

- **<salute>** marks a salutation at either the beginning or the end of a text. This goes within the <body> tag of a text, and is generally a part of an <opener> or a <closer>. This tag cannot be placed within a <p> or a <note>.

- **<signed>** indicates the signature of the writer. This tag goes within the <body> tag, generally as a part of the <closer>. It cannot be put inside of a <p> or a <note>.

- **<address>** surrounds an entire address in any format.

- **<adrLine>** is a sub-element of <address> that signifies each individual line of an address.

4.3.3 Tags Used for Future Markup of Transcriptions

The following tags, although already included in the tag set, will not be used in during the initial transcriptions of the Robinson Reports. They may, however, be added to the transcriptions at a later date to more explicitly mark up the text.

These tags will not be included during the initial transcription for several reasons. First, Professor Fontanella will be on a very tight schedule while he is in London, which will limit the amount of tagging that he can do. He also may not have the knowledge to address certain issues, such as the regularised cataloguing of names. Finally, there are some concepts that have not been prepared for inclusion at this time, such as digitised images of the reports or pictures within them.

It would be very beneficial to the NAL and art scholars using the system to add these tags to the transcriptions after Fontanella finishes his work. For this purpose, more detailed explanations of these tags, along with examples, have been included in the *Guidelines for Developing the Search System* included as Appendix G.

4.3.3.1 Components of Names of People

These tags are used for breaking up people's names into their smaller components. This would be very beneficial for cataloguing purposes, as it would facilitate searching for names by allowing users to specify the part of the name they are

looking for. For instance, a general name search for "Lawrence" would bring up a lot more extraneous information than a search for "Lawrence" as a surname. All of the elements in this section are subelements of <persName>. Additionally, splitting up names into their components eases the process of formulating alphebetised or otherwise ordered lists.

- **<surname>** marks any surnames given within the <persName> tag.
- **<forename>** indicates any forenames given.
- **<rolename>** signifies any information pertaining to the person's role. This could include a rank or title.
- **<addName>** is used to mark a nickname, epithet, alias, or other phrase that is not part of the person's actual name.
- **<nameLink>** contains a linking phrase in a name, such as "di" or "van der."
- **<genName>** is used to tag generational information, such as "junior" or "the third."

4.3.3.2 Components of Names of Places

Like the names of people, the names of places can be split up into smaller pieces. Different levels of geo-political units, as well as geographical features, can be specified within the <placeName> tag. All of the elements in this section are subelements of <placeName>.

- **<settlement>** denotes the smallest existing geo-political unit, such as a city or town.
- **<region>** indicates a geo-political unit larger than settlement, but smaller than country. For example, a state or province would be a region.
- **<country>** is used to tag the name of a country.
- **<bloc>** indicates a geo-political unit containing one or more countries.

- **<geogName>** signifies the name of a specific geographical feature, such as "Mount Everest" or "Death Valley."

- **<geog>** is used to mark a common noun representing some geographical feature, such as "mount" or "valley."

- **<distance>** can be a sub-element of <placeName>, <dateStruct>, or <timeStruct>. It denotes the amount of space or time between two things.

- **<offset>** can be a sub-element of <placeName>, <dateStruct>, or <timeStruct>. It is used in conjunction with <distance> to specify direction.

4.3.3.3 Components of Names of Organisations

Names of organisations can also be broken up into smaller subelements. This allows specific parts of an organisation to be noted, while still including it as a part of the larger organisation. In addition, specifying the type of organisation allows future analysts to generate lists of certain types of organisations mentioned within the text.

- **<orgtitle>** is a sub-element of <orgName> that denotes the proper title of the organisation.

- **<orgtype>** is a sub-element of <orgName> that is used to mark information pertaining to the type of organisation being named.

- **<orgdivn>** is a sub-element of <orgName>, which indicates a branch or division of the organisation.

4.3.3.4 Inserting Images

The two tags described in this section are used for inserting pointers to images into the text. This would allow the software displaying the document to find that image and display it with the text. Some of the Robinson Reports include sketches of the objects being discussed. If these sketches were digitised, the electronic edition of the Robinson Reports could include them using the tags described below.

- **<figure>** is used to create a link from the text to an image file.

- **<figDesc>** is used in conjunction with <figure> to insert a description of the ssimage.

4.3.3.5 Linking

The tags in this section are used for creating links between different areas of the text. Links are made by referring to the "id" attribute of an element; to facilitate linking, tags are provided that mark otherwise unmarked areas of the text. Linking should be used to provide hyperlinks to related letters, as well as to Fontanella's comments and interpretation. However, linking is a complex process that requires a unique id for each point of reference. Because of this, it should be done after the transcriptions are complete. For information on other methods of segmentation and linking, see Appendix E.

- **<seg>** is used to designate a segment of text that has not been otherwise tagged. This segment can then have interpretations or comments linked to it.

- **<anchor>** is used to mark a particular spot in a text, without marking any of the text itself.

- **<link>** provides pointers between two elements. For instance, a <seg> tag can be linked to a <note> tag. This would create a pointer from the <seg> tag to the <note> tag and from the <note> tag to the <seg> tag. This could be useful if the note contains an interpretation of the segment of text.

- **<linkGrp>** is used to group a series of links together. <linkGrp> is useful because it allows attributes to be specified for the entire group; hence, they do not have to be entered for each link.

- **<xptr>** allows for pointers to be set to outside sources. This could be used for inserting digitised pictures or references to outside documents.

4.4 Transcription Templates

Based on the document types described in Table 2 in Section 4.1.1, we developed twelve templates to facilitate Professor Fontanella's transcriptions. These templates were

developed for WordPerfect 9, and conform to the DTD that we specified for the transcription of the Robinson Reports. These templates define the structure of the various types of documents and include pre-printed text that was present on standard forms. They also provide shortcut keys and buttons to the common phrases that we drew from our analysis of the Robinson Reports, found in Table 4 in Section 4.1.3. Images of the templates, as well as explicit documentation on their use, are available in the *User Manual for Transcribing the Robinson Reports* included as Appendix F. Electronic copies of the templates are also available on the enclosed CD.

4.5 Sample Transcriptions

The document templates that we developed were used in the creation of several sample transcriptions. We wrote some of these sample transcriptions, while others were written by Professors Demetry and Vaz of WPI. These transcriptions have been included with the templates in the *User Manual for Transcribing the Robinson Reports* in Appendix F and in an electronic form on the enclosed CD. This will help Professor Fontanella visualise the final transcriptions better, and should make it easier for him to understand and use the templates. Replicas of the letters used in the sample transcriptions have also been included in the user manual. This will allow Fontanella to practice using the software before arriving in London by replicating the transcriptions that are already completed.

4.6 Estimates of Transcription Time

It was estimated that Professor Fontanella will have fourteen minutes in the NAL to work on each page of the Robinson Reports. This was based on an estimate of 3500

pages in the collection, with Professor Fontanella staying in London for seventy-five days and working eight hours a day for five and a half days each week.

The sample transcriptions completed by Professors Demetry and Vaz, along with those that we completed ourselves, took an average of forty minutes to fully transcribe and mark up; however, these numbers are not as bad as they appear. The XML editor and tagging system are quite complicated to the new user. It was very apparent that there was a learning curve that affected the time necessary to transcribe each report. Although users took nearly an hour to complete their first transcriptions, by the third or fourth transcription they were taking only twenty to twenty-five minutes.

If Fontanella studies the user manual and practices with the sample letters before beginning his work in London, he should be able to complete the transcriptions in twenty minutes at the most. Furthermore, the transcriptions do not have to be entirely completed in the NAL. Fontanella's comments and interpretations, as well as many of the tags that describe the content, can be added to the transcriptions after the library closes. If this is done, he should be able to finish each page within the fourteen minute time frame.

5 Prospects and Recommendations

The transcription and encoding of the Robinson Reports will be beneficial to both the NAL and scholars. When the reports are made available electronically, scholars will be able to make use of them in their research. Because the transcriptions will be encoded in a standard non-proprietary format, they will be easily interpreted by many software applications, which will make them accessible to a wider range of scholars and organisations. In addition, the work done with the Robinson Reports will advance future research into standard formats for transcribing and encoding documents. As more collections are made available electronically, the range and efficiency of scholarly research will be greatly increased.

After the encoding of the Robinson Reports, a system will be developed for searching the transcriptions. Most modern cataloguing systems use keywords to search by subject. The system that will be developed for the Robinson Reports will search via tags embedded in the text of the transcriptions, which is much more efficient and accurate than searching by keyword.

The search system will aid research because it will allow scholars to use specific criteria to locate the transcriptions that are beneficial to their research. They will no longer have to painstakingly look through the reports searching for pertinent information. The system will allow scholars to search the reports using a variety of criteria, including names of people, geographic locations, and pieces of art. For example, if scholars are looking for all of the reports discussing pieces from Italy, they will be able to search for all the transcriptions that have the word "Italy" tagged as the name of a place. In seconds

the system will be able to locate and display all of the transcriptions that meet this criterion.

To facilitate the completion of the search system, we documented our recommendations for its development. These recommendations include suggestions for optimising the markup of the transcriptions for searching and displaying, instructions for combining the documents to reduce the number of files, the search criteria that should be provided to the user, and descriptions of how the transcriptions should be interpreted so that style sheets can be correctly implemented. These *Guidelines for Completing the System* can be found in Appendix G.

The *Guidelines for Completing the System* could be used in the implementation of future projects for the NAL and other organisations. The tag set we developed is extremely broad and can be used to mark up a wide variety of documents. The NAL and any other interested organisations will be able to use our tag set to encode transcriptions of other collections. If our tag set is insufficient for the collection they wish to encode, they have the option to create a new tag set based on ours. TEI provides many additional tag sets that are not included in ours because we felt they were not applicable to the Robinson Reports. Furthermore, the TEI standard is still growing, with new tags and tag sets constantly being developed. New and existing tag sets can be added to our's for the transcription of other documents. For example, TEI provides an additional tag set specifically designed for encoding poetry. This tag set would be useless to the Robinson Reports; however, it would be a very useful addition for transcribing an epic poem.

The work being done with the Robinson Reports will provide a useful tool for scholars interested in museums, art, and particularly the history of how the V&A's

collection was built and developed. More importantly, however, it will provide a stepping stone for future work with standard information interchange. Continued work in this area will eventually lead to the availability of a multitude of sources that were previously inaccessible, which will benefit scholars and curious readers alike.

6 Bibliography

Baker, Malcolm, and Brenda Richardson, Eds. A Grand Design: The Art of the Victoria and Albert Museum. New York: Harry N Abrams, 1997.

Breuck, Lora. Personal interview. 18 Feb. 2000.

Burns, Jill. Personal interview. 11 Feb. 2000

CIMI Home. 27 March 2000. Consortium for the Computer Interchange of Museum Information. Accessed 2 May 2000. <http://www.cimi.org>.

Degutis, Alan. Personal interview. 11 Feb. 2000

Fontanella, Lee. Charles Thurston Thompson and the Iberian Photographic Project.

Fontanella, Lee. Personal interview. Jan/Feb. 2000.

Furrie, Bettie. Understanding MARC Bibliographic: Machine-Readable Cataloguing. 1998. Library of Congress. Accessed 25 Jan. 2000 <http://lcweb.loc.gov/marc/umb>.

Hachem, Nabil. Personal interview. 8 Feb. 2000

Heaps, H. S. Information Retrieval: Computational and Theoretical Aspects. New York: Academic Press, 1978.

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

Holley, Robert P., ed. Subject Control in Online Catalogues. New York: Haworth Press, 1989.

Johnson, Mark. XML for the Absolute Beginner. April 1999. JavaWorld. Accessed 27 Feb. 2000 <http:/ /www.javaworld.com/jw-04-1999/f_jw-04-xml_p.html>.

Lancaster, Frederick Wilfrid. Information Retrieval Systems: Characteristics, Testing, and Evaluation. New York: John Wiley & Sons, 1979.

McDowell, Earl E. Interviewing Practices for Technical Writers. Amityville, NY: Baywood, 1991.

McRae, Linda, and Lynda S. White, Eds. ArtMARC Sourcebook: Cataloguing Art, Architecture and Their Visual Images. Chicago: American Library Association,

1998.

Meadow, Charles T. The Analysis of Information Systems. Los Angeles: Melville, 1973.

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

Rowland, Arthur Ray, Ed. The Catalog and Cataloging. Shoe String, 1969.

Sir John Charles Robinson - Biography. National Gallery of Art. Accessed 19 Jan. 2000 <http://www.nga.gov/cgi-bin/pbio?230780>.

TEI Guidelines for Electronic Text Encoding and Interchange (P3). 16 May 1994. Electronic Text Center of the University of Virginia. Accessed 28 March 2000 <http://etext.virginia.edu/TEI.html>.

Text Encoding Initiative Home Page. 28 June 1999. Text Encoding Initiative. Accessed 2 May 2000. <http://www.uic.edu/orgs/tei/lite/index.html>.

V & A. Home page. 20 Jan. 2000. Victoria and Albert Museum. Accessed 23 Jan. 2000 <http://www.vam.ac.uk/Infodome/>.

Wait, Brad. Using XML in Oracle Database Applications Part 1: About XML. Nov. 1999. Oracle. Accessed 2 Feb. 2000 <http://technet.oracle.com/tech/xml/info/htdocs/otnwp/about_xml.htm>.

Wynar, Bohdan S. Introduction to Cataloguing and Classification. 3rd ed. Rochester: Libraries Unlimited, 1967.

7 Appendix A – The Victoria and Albert Museum

The Victoria and Albert Museum (V&A) was founded in 1852 and was originally named the Museum of Manufactures (Baker 13). Henry Cole, the original director, felt that this name did not complement the historical collection that he was building so he changed the name in 1853 to a more historical-sounding name, the Museum of Ornamental Art (Baker 151). The museum resided in the Marlborough House in the centre of Kensington until it was moved in 1857 to its current residence, the Brompton fields in South Kensington. Upon the move, the museum was renamed as the South Kensington Museum. Finally, in 1899 it attained its current name, the Victoria and Albert Museum, in honour of Queen Victoria and her deceased husband Prince Albert (Baker 13).

The V&A is one of the most prestigious museums in the world. In fact, it is the world's largest decorative arts museum. The museum holds England's national collections of furniture, sculpture, glass, ceramics, watercolours, portrait miniatures, photographs, and its National Art Library. The museum has 146 galleries and over 4 million items. These items range from the Constable paintings to one of the premiere collections of Italian Renaissance sculpture (V&A website).

From its outset, the goal of the V&A has been to educate the general public about the decorative arts. To sustain its goal the V&A has set up circulating collections of artwork with museums around the world. The museum has also stressed the point of explaining the art rather than just displaying it. As a result of its great success the V&A has become one of the pioneering museums in the world. It has set standards by which all museums have set out to abide (Baker 9-13).

Along with the main museum in South Kensington, the V&A is also responsible for the Bethnal Green Museum of Childhood, the Wellington Museum at Apsley House and the Theatre Museum in Covent Garden (V&A website).

8 Appendix B – Examples of the Robinson reports

Professor Fontanella's book *Charles Thurston Thompson* contains several excerpts from the Robinson Reports. These excerpts have been included here to provide an idea of the type of information the letters contain. It is important to note, however, that these letters all pertain to Fontanella's research regarding Charles Thurston Thompson, the V&A's photographer at the time, and do not accurately represent the extreme diversity of the content and style of the Robinson Reports. It is also important to mention that these are excerpts and not full transcriptions of letters.

8.1 30 May 1866

> I am more than ever convinced that this country is still rich in works of art, hidden in all parts of the great unexplored territory, but a great expenditure of time and patience is requisite in order to get at them. My time has not been ill spent in Madrid, and I do not regret my continued stay here, though it has been personally a very disagreeable one.
>
> [...]
>
> For all negotiations unfortunately an interminable time is required, all [?] alike in Spain are endowed with a force of inertia and innate and habitual slowness, which there is no alternative but to fall in with. On the other hand the most complete ignorance both of the value and merit of work of art prevails, and it is evident to me that it is here, i.e. in Spain alone, in the present day, that works of art of real importance remain still to be discovered (Fontanella 39).

8.2 6 September 1866

> the conditions both in France and the north of Spain have intimidated communication in all directions. [...] this country is in semi revolution, money has disappeared, [...] O'Donnell [Leopold O'Donnell y Jorris, Count of Lucena and Duke of Tetuán 1809-1867] has confiscated one fourth of the year's salary of all the clergy in Spain, & as a consequence they will sell their [?] cassocks & mitres even [...] & anything they have in the way of works of art, or anything they can make a shilling by will soon come to the surface (Fontanella 39).

8.3 23 October 1866

The independence of the Cathedral authorities is such that it would not be sufficient to apply to them through direct official channels (though this *would* be requisite in the form of an application from our Minister in Madrid to the Spanish Minister "de Gracias y Cultos"). There exists, however, another authoritative body in Madrid, which could probably surmount all obstacles; this is the "Real Academia de San Fernando." This academy has heretofore been a very august institution, with little better than nominal functions, the academicians being mainly great personages of the court, artists being in a small minority. Latterly, however, efforts have been made to reanimate this institution, and it has recently acquired a kind of semi-official supervision over all the ancient monuments and public works of art in the country. Local corresponding members and committees are also being appointed in the provincial cities (Fontanella 34).

8.4 18 December 1866

Generally speaking the interiors of the architectural monuments of Spain are *everything* and the exteriors of little moment [?] —moreover a great number of *picturesque exterior views* have already been taken by photographers resident in Spain, and they are being multiplied every day. I do not think therefore it would be necessary to keep Mr. Thompson in Spain for *out door* photographs.

But with regard to *the interior subjects* the case is difficult —I cannot overrate the importance of proceeding [?][...]ally or procuring photos of the chef d'oeuvres of sculpture, ornamentation church[...][?], which abounds in every ancient city of the Peninsula. The only point to establish is whether under this condition of semi-darkness which prevails in almost all Spanish churches, *it is possible by any means to produce satisfactory photographs*.

When I have seen Mr. Thompson's photographs from Santiago, Coimbra, & Batalha, I shall be able to advice[sic] with more certainty as to his future operations (Fontanella 36).

9 Appendix C – Interview Questions

Throughout our research we conducted several interviews to aid our work (see Chapter 3.2). These interviews included Professors Fontanella and Hachem of WPI, Alan Degutis of the American Antiquarian Society, Jill Burns of the Worcester Art Museum, Lora Breuck of the WPI library, and Cathy Peterson, a web application developer. Although we ran our interviews as open-ended discussions, rather than question and answer sessions, we did develop specific questions. These questions, included here, were used as references to keep our interviews on track and make sure that we covered all the necessary information.

9.1 Interview of Professor Fontanella

The following questions were used in our interview of Professor Fontanella on 25 January 2000.

> What purpose did the Robinson reports serve and what specific information is contained within them?
>
> Is the category info, which is referred to in the project specifications, in the letters or is it provided to us by some other means?
>
> What will be stored in the system, the direct transcriptions, just the info listed in the proposal, or both?
>
> Are the Robinson reports from the Victoria and Albert Museum going to be included in this system as well? (Refer to the proposal where it mentions only the microfilm being added to the system)
>
> In an ideal world what would you want the system to do for you and what do you think other scholars would want from the system? What are the challenges that this system has to overcome?

Are we permitted to have software requirements for the system? For example, if we were to use Access for the system, is it all right to require that the user have Access to run the system?

About how many pieces of art need to be catalogued?

Do you have any written documentation about this subject?

Do you know of any other people or sources we could get some more information from?

9.2 Interviews at Libraries and Museums

The following questions were used in our interviews at libraries and museums. These include the interviews of Alan Degutis and Jill Burns on 11 February 2000 and Lora Breuck on 18 February 2000.

What software is used for your cataloguing system?

What is the overall setup for the system?

How do you handle data entry (authorised) vs. data retrieval (public)? Are there separate interfaces?

Do you have any user feedback on your system?

What method do you use to uniquely identify each item in the collection?

In what way do you categorise the items in your collection?

10 Appendix D – The TEI Extension Files

VAExt.ent:

```
<!-- * VAExt.ent
     *
     * This extension file defines the extensions used in combination
with the
     * VAExt.dtd
     *
     * Version 1
     * 19 April 2000
     * -->

<!ENTITY % x.phrase 'collectionName | eventName | materialName
                     | objectName | titleName | objectIdentifier | cost
|'  >

<!ENTITY % a.terminology '
          class                  CDATA               #IMPLIED'      >

<!ENTITY % name                        'IGNORE' >
<!ENTITY % rs                          'IGNORE' >
<!ENTITY % measure                     'IGNORE' >
<!ENTITY % pubPlace                    'IGNORE' >
<!ENTITY % persName                    'IGNORE' >
<!ENTITY % surname                     'IGNORE' >
<!ENTITY % forename                    'IGNORE' >
<!ENTITY % genName                     'IGNORE' >
<!ENTITY % nameLink                    'IGNORE' >
<!ENTITY % addName                     'IGNORE' >
<!ENTITY % placeName                   'IGNORE' >
<!ENTITY % settlement                  'IGNORE' >
<!ENTITY % region                      'IGNORE' >
<!ENTITY % country                     'IGNORE' >
<!ENTITY % bloc                        'IGNORE' >
<!ENTITY % geogName                    'IGNORE' >
<!ENTITY % orgName                     'IGNORE' >
```

VAExt.dtd:

```
<!-- * VAExt.dtd
     *
     * This extension file defines the CIMI tags and modifcations that
are to be
     * merged into the TEI tag set generated for the Victoria and
Albert Museum
     *
     * Version 1
     * 19 April 2000
     *
-->

<!-- Certain Portions of this document were copied from the TEI
document files-->
<!-- Text Encoding and Interchange. Document TEI P3, 1994.
-->
<!-- Copyright (c) 1994 ACH, ACL, ALLC. Permission to copy
-->
<!-- in any form is granted, provided this notice is
-->
<!-- included in all copies.
-->

<!-- Added CIMI tags -->

<!ELEMENT collectionName     ((#PCDATA | note)*)                >
<!ATTLIST collectionName     %a.global;
                             %a.names;
          authority          (Non-Authoritative | Local-to-Server | AAT
| AAT.Date |
                             ACRL.RBMS-Binding | ACRL.RBMS-Genre |
ACRL.RBMS-Paper | ACRL.RBMS-Printing
                             | ACRL.RBMS-Type | Base-Merimee | BGN |
British-Archaelogical |
                             Canadiana | CDWA | Dictionarium-
Museologicu | Garnier | Geosaurus
                             | Glass | ICOM.Costume | Iconclass |
Jewish-Art | ISO.Language |
                             ISO.Documentation | ISO.Iconic | ISO.AV |
ISO.Date-Time | LC.Descriptive-Graphic
                             | LC.Name | LC.Thesaurus-Graphic | LCSH |
Moving-Image-Materials |
                             Nomenclature | Reynies | TGN | Tozzer |
Ulan | USMarc | Villard
                             | Yale.British-Artists)  #IMPLIED
          role               CDATA               #IMPLIED
          type               CDATA               #IMPLIED
          TEIform            CDATA               'collectionName'>


<!ELEMENT eventName          ((#PCDATA | note)*)                >
<!ATTLIST eventName          %a.global;
                             %a.names;
```

```
          authority          (Non-Authoritative | Local-to-Server | AAT
| AAT.Date |
                             ACRL.RBMS-Binding | ACRL.RBMS-Genre |
ACRL.RBMS-Paper | ACRL.RBMS-Printing
                             | ACRL.RBMS-Type | Base-Merimee | BGN |
British-Archaelogical |
                             Canadiana | CDWA | Dictionarium-
Museologicu | Garnier | Geosaurus
                             | Glass | ICOM.Costume | Iconclass |
Jewish-Art | ISO.Language |
                             ISO.Documentation | ISO.Iconic | ISO.AV |
ISO.Date-Time | LC.Descriptive-Graphic
                             | LC.Name | LC.Thesaurus-Graphic | LCSH |
Moving-Image-Materials |
                             Nomenclature | Reynies | TGN | Tozzer |
Ulan | USMarc | Villard
                             | Yale.British-Artists)  #IMPLIED
          role               CDATA               #IMPLIED
          type               CDATA               #IMPLIED
          TEIform            CDATA               'eventName'    >


<!ELEMENT materialName       ((#PCDATA | note)*)                >
<!ATTLIST materialName       %a.global;
                             %a.names;
          authority          (Non-Authoritative | Local-to-Server | AAT
| AAT.Date |
                             ACRL.RBMS-Binding | ACRL.RBMS-Genre |
ACRL.RBMS-Paper | ACRL.RBMS-Printing
                             | ACRL.RBMS-Type | Base-Merimee | BGN |
British-Archaelogical |
                             Canadiana | CDWA | Dictionarium-
Museologicu | Garnier | Geosaurus
                             | Glass | ICOM.Costume | Iconclass |
Jewish-Art | ISO.Language |
                             ISO.Documentation | ISO.Iconic | ISO.AV |
ISO.Date-Time | LC.Descriptive-Graphic
                             | LC.Name | LC.Thesaurus-Graphic | LCSH |
Moving-Image-Materials |
                             Nomenclature | Reynies | TGN | Tozzer |
Ulan | USMarc | Villard
                             | Yale.British-Artists)  #IMPLIED
          role               CDATA               #IMPLIED
          type               CDATA               #IMPLIED
          TEIform            CDATA               'materialName' >


<!ELEMENT objectName         ((#PCDATA | note)*)                >
<!ATTLIST objectName         %a.global;
                             %a.names;
          authority          (Non-Authoritative | Local-to-Server | AAT
| AAT.Date |
                             ACRL.RBMS-Binding | ACRL.RBMS-Genre |
ACRL.RBMS-Paper | ACRL.RBMS-Printing
                             | ACRL.RBMS-Type | Base-Merimee | BGN |
British-Archaelogical |
```

```
                              Canadiana | CDWA | Dictionarium-
Museologicu | Garnier | Geosaurus
                              | Glass | ICOM.Costume | Iconclass |
Jewish-Art | ISO.Language |
                              ISO.Documentation | ISO.Iconic | ISO.AV |
ISO.Date-Time | LC.Descriptive-Graphic
                              | LC.Name | LC.Thesaurus-Graphic | LCSH |
Moving-Image-Materials |
                              Nomenclature | Reynies | TGN | Tozzer |
Ulan | USMarc | Villard
                              | Yale.British-Artists)  #IMPLIED
          role                CDATA                 #IMPLIED
          type                CDATA                 #IMPLIED
          TEIform             CDATA                 'objectName'   >


<!ELEMENT titleName           (%paraContent;)                      >
<!ATTLIST titleName           %a.global;
                              %a.names;
          authority           (Non-Authoritative | Local-to-Server | AAT
| AAT.Date |
                              ACRL.RBMS-Binding | ACRL.RBMS-Genre |
ACRL.RBMS-Paper | ACRL.RBMS-Printing
                              | ACRL.RBMS-Type | Base-Merimee | BGN |
British-Archaelogical |
                              Canadiana | CDWA | Dictionarium-
Museologicu | Garnier | Geosaurus
                              | Glass | ICOM.Costume | Iconclass |
Jewish-Art | ISO.Language |
                              ISO.Documentation | ISO.Iconic | ISO.AV |
ISO.Date-Time | LC.Descriptive-Graphic
                              | LC.Name | LC.Thesaurus-Graphic | LCSH |
Moving-Image-Materials |
                              Nomenclature | Reynies | TGN | Tozzer |
Ulan | USMarc | Villard
                              | Yale.British-Artists)  #IMPLIED
          role                CDATA                 #IMPLIED
          type                CDATA                 #IMPLIED
          TEIform             CDATA                 'titleName'    >


<!ELEMENT objectIdentifier    ((#PCDATA | note)*)                        >
<!ATTLIST objectIdentifier    %a.global;
                              %a.names;
          authority           (Non-Authoritative | Local-to-Server | AAT
| AAT.Date |
                              ACRL.RBMS-Binding | ACRL.RBMS-Genre |
ACRL.RBMS-Paper | ACRL.RBMS-Printing
                              | ACRL.RBMS-Type | Base-Merimee | BGN |
British-Archaelogical |
                              Canadiana | CDWA | Dictionarium-
Museologicu | Garnier | Geosaurus
                              | Glass | ICOM.Costume | Iconclass |
Jewish-Art | ISO.Language |
                              ISO.Documentation | ISO.Iconic | ISO.AV |
ISO.Date-Time | LC.Descriptive-Graphic
```

```
                              | LC.Name | LC.Thesaurus-Graphic | LCSH |
Moving-Image-Materials |
                              Nomenclature | Reynies | TGN | Tozzer |
Ulan | USMarc | Villard
                              | Yale.British-Artists)  #IMPLIED
          role                CDATA                 #IMPLIED
          type                CDATA                 #IMPLIED
          TEIform             CDATA                 'objectIdentifier'   >


<!ELEMENT cost                ((#PCDATA | note)*)                        >
<!ATTLIST cost                %a.global;
          TEIform             CDATA                 'cost'               >

<!-- Modifications to existing TEI tags -->

<!-- Added the following to name tag:
 ELEMENTS
   collectionName
   eventName
   materialName
   objectName
   titleName
   objectIdentifier

 ATTRIBUTES
   authority
   role
-->

<!ELEMENT name                (%phrase.seq;)                          >
<!ATTLIST name                %a.global;
                              %a.names;
          authority           (Non-Authoritative | Local-to-Server | AAT
| AAT.Date |
                              ACRL.RBMS-Binding | ACRL.RBMS-Genre |
ACRL.RBMS-Paper | ACRL.RBMS-Printing
                              | ACRL.RBMS-Type | Base-Merimee | BGN |
British-Archaelogical |
                              Canadiana | CDWA | Dictionarium-
Museologicu | Garnier | Geosaurus
                              | Glass | ICOM.Costume | Iconclass |
Jewish-Art | ISO.Language |
                              ISO.Documentation | ISO.Iconic | ISO.AV |
ISO.Date-Time | LC.Descriptive-Graphic
                              | LC.Name | LC.Thesaurus-Graphic | LCSH |
Moving-Image-Materials |
                              Nomenclature | Reynies | TGN | Tozzer |
Ulan | USMarc | Villard
                              | Yale.British-Artists)  #IMPLIED
          role                CDATA                 #IMPLIED
          type                CDATA                 #IMPLIED
          TEIform             CDATA                 'name'          >


<!-- Added the following to rs tag:
 ELEMENTS
```

```
   collectionName
   eventName
   materialName
   objectName
   titleName
   objectIdentifier

 ATTRIBUTES
   authority
   role
-->

<!ELEMENT rs                  (%phrase.seq;)                      >
<!ATTLIST rs                  %a.global;
                              %a.names;
          authority           (Non-Authoritative | Local-to-Server | AAT
| AAT.Date |
                              ACRL.RBMS-Binding | ACRL.RBMS-Genre |
ACRL.RBMS-Paper | ACRL.RBMS-Printing
                              | ACRL.RBMS-Type | Base-Merimee | BGN |
British-Archaelogical |
                              Canadiana | CDWA | Dictionarium-
Museologicu | Garnier | Geosaurus
                              | Glass | ICOM.Costume | Iconclass |
Jewish-Art | ISO.Language |
                              ISO.Documentation | ISO.Iconic | ISO.AV |
ISO.Date-Time | LC.Descriptive-Graphic
                              | LC.Name | LC.Thesaurus-Graphic | LCSH |
Moving-Image-Materials |
                              Nomenclature | Reynies | TGN | Tozzer |
Ulan | USMarc | Villard
                              | Yale.British-Artists)  #IMPLIED
          role                CDATA                #IMPLIED
          type                CDATA                #IMPLIED
          TEIform             CDATA                'rs'           >


<!-- Added the following to measure tag:
 ELEMENTS
   collectionName
   eventName
   materialName
   objectName
   titleName
   objectIdentifier

 ATTRIBUTES
   authority
   role
   extent
   value
   unit
-->

<!ELEMENT measure             (%phrase.seq;)                      >
<!ATTLIST measure             %a.global;
                              %a.names;
```

```
          authority           (Non-Authoritative | Local-to-Server | AAT
| AAT.Date |
                              ACRL.RBMS-Binding | ACRL.RBMS-Genre |
ACRL.RBMS-Paper | ACRL.RBMS-Printing
                              | ACRL.RBMS-Type | Base-Merimee | BGN |
British-Archaelogical |
                              Canadiana | CDWA | Dictionarium-
Museologicu | Garnier | Geosaurus
                              | Glass | ICOM.Costume | Iconclass |
Jewish-Art | ISO.Language |
                              ISO.Documentation | ISO.Iconic | ISO.AV |
ISO.Date-Time | LC.Descriptive-Graphic
                              | LC.Name | LC.Thesaurus-Graphic | LCSH |
Moving-Image-Materials |
                              Nomenclature | Reynies | TGN | Tozzer |
Ulan | USMarc | Villard
                              | Yale.British-Artists)  #IMPLIED
          role                CDATA                #IMPLIED
          type                CDATA                #IMPLIED
          extent              CDATA                #IMPLIED
          value               CDATA                #IMPLIED
          unit                CDATA                #IMPLIED
          TEIform             CDATA                'measure'      >


<!-- Added the following to pubPlace tag:
 ELEMENTS
   collectionName
   eventName
   materialName
   objectName
   titleName
   objectIdentifier

 ATTRIBUTES
   authority
   role
-->

<!ELEMENT pubPlace            (%phrase.seq;)                      >
<!ATTLIST pubPlace            %a.global;
                              %a.names;
          authority           (Non-Authoritative | Local-to-Server | AAT
| AAT.Date |
                              ACRL.RBMS-Binding | ACRL.RBMS-Genre |
ACRL.RBMS-Paper | ACRL.RBMS-Printing
                              | ACRL.RBMS-Type | Base-Merimee | BGN |
British-Archaelogical |
                              Canadiana | CDWA | Dictionarium-
Museologicu | Garnier | Geosaurus
                              | Glass | ICOM.Costume | Iconclass |
Jewish-Art | ISO.Language |
                              ISO.Documentation | ISO.Iconic | ISO.AV |
ISO.Date-Time | LC.Descriptive-Graphic
                              | LC.Name | LC.Thesaurus-Graphic | LCSH |
Moving-Image-Materials |
```

```
                              Nomenclature | Reynies | TGN | Tozzer |
Ulan | USMarc | Villard
                              | Yale.British-Artists)  #IMPLIED
          role                CDATA                  #IMPLIED
          TEIform             CDATA                  'pubPlace'     >


<!-- Added the following to persName tag:
 ELEMENTS
   collectionName
   eventName
   materialName
   objectName
   titleName
   objectIdentifier

 ATTRIBUTES
   authority
   role
-->

<!ELEMENT persName            (#PCDATA | %m.personPart; |
                              %m.phrase;)*                        >
<!ATTLIST persName            %a.global;
                              %a.names;
          authority           (Non-Authoritative | Local-to-Server | AAT
| AAT.Date |
                              ACRL.RBMS-Binding | ACRL.RBMS-Genre |
ACRL.RBMS-Paper | ACRL.RBMS-Printing
                              | ACRL.RBMS-Type | Base-Merimee | BGN |
British-Archaelogical |
                              Canadiana | CDWA | Dictionarium-
Museologicu | Garnier | Geosaurus
                              | Glass | ICOM.Costume | Iconclass |
Jewish-Art | ISO.Language |
                              ISO.Documentation | ISO.Iconic | ISO.AV |
ISO.Date-Time | LC.Descriptive-Graphic
                              | LC.Name | LC.Thesaurus-Graphic | LCSH |
Moving-Image-Materials |
                              Nomenclature | Reynies | TGN | Tozzer |
Ulan | USMarc | Villard
                              | Yale.British-Artists)  #IMPLIED
          role                CDATA                  #IMPLIED
          type                CDATA                  #IMPLIED
          TEIform             CDATA                  'persName'     >


<!-- Added the following to surname tag:
 ELEMENTS
   collectionName
   eventName
   materialName
   objectName
   titleName
   objectIdentifier

 ATTRIBUTES
```

```
   authority
   role
-->

<!ELEMENT surname            (%phrase.seq;)                    >
<!ATTLIST surname            %a.global;
                             %a.personPart;
          authority          (Non-Authoritative | Local-to-Server | AAT
| AAT.Date |
                             ACRL.RBMS-Binding | ACRL.RBMS-Genre |
ACRL.RBMS-Paper | ACRL.RBMS-Printing
                             | ACRL.RBMS-Type | Base-Merimee | BGN |
British-Archaelogical |
                             Canadiana | CDWA | Dictionarium-
Museologicu | Garnier | Geosaurus
                             | Glass | ICOM.Costume | Iconclass |
Jewish-Art | ISO.Language |
                             ISO.Documentation | ISO.Iconic | ISO.AV |
ISO.Date-Time | LC.Descriptive-Graphic
                             | LC.Name | LC.Thesaurus-Graphic | LCSH |
Moving-Image-Materials |
                             Nomenclature | Reynies | TGN | Tozzer |
Ulan | USMarc | Villard
                             | Yale.British-Artists)  #IMPLIED
          role               CDATA               #IMPLIED
          TEIform            CDATA               'surname'      >


<!-- Added the following to forename tag:
 ELEMENTS
   collectionName
   eventName
   materialName
   objectName
   titleName
   objectIdentifier

 ATTRIBUTES
   authority
   role
-->

<!ELEMENT forename           (%phrase.seq;)                     >
<!ATTLIST forename           %a.global;
                             %a.personPart;
          authority          (Non-Authoritative | Local-to-Server | AAT
| AAT.Date |
                             ACRL.RBMS-Binding | ACRL.RBMS-Genre |
ACRL.RBMS-Paper | ACRL.RBMS-Printing
                             | ACRL.RBMS-Type | Base-Merimee | BGN |
British-Archaelogical |
                             Canadiana | CDWA | Dictionarium-
Museologicu | Garnier | Geosaurus
                             | Glass | ICOM.Costume | Iconclass |
Jewish-Art | ISO.Language |
                             ISO.Documentation | ISO.Iconic | ISO.AV |
ISO.Date-Time | LC.Descriptive-Graphic
```

```
                              | LC.Name | LC.Thesaurus-Graphic | LCSH |
Moving-Image-Materials |
                              Nomenclature | Reynies | TGN | Tozzer |
Ulan | USMarc | Villard
                              | Yale.British-Artists)  #IMPLIED
          role                CDATA                 #IMPLIED
          TEIform             CDATA                 'forename'      >


<!-- Added the following to genName tag:
 ELEMENTS
   collectionName
   eventName
   materialName
   objectName
   titleName
   objectIdentifier

 ATTRIBUTES
   authority
   role
-->

<!ELEMENT genName             (%phrase.seq;)                        >
<!ATTLIST genName             %a.global;
                              %a.personPart;
          authority           (Non-Authoritative | Local-to-Server | AAT
| AAT.Date |
                              ACRL.RBMS-Binding | ACRL.RBMS-Genre |
ACRL.RBMS-Paper | ACRL.RBMS-Printing
                              | ACRL.RBMS-Type | Base-Merimee | BGN |
British-Archaelogical |
                              Canadiana | CDWA | Dictionarium-
Museologicu | Garnier | Geosaurus
                              | Glass | ICOM.Costume | Iconclass |
Jewish-Art | ISO.Language |
                              ISO.Documentation | ISO.Iconic | ISO.AV |
ISO.Date-Time | LC.Descriptive-Graphic
                              | LC.Name | LC.Thesaurus-Graphic | LCSH |
Moving-Image-Materials |
                              Nomenclature | Reynies | TGN | Tozzer |
Ulan | USMarc | Villard
                              | Yale.British-Artists)  #IMPLIED
          role                CDATA                 #IMPLIED
          TEIform             CDATA                 'genName'       >


<!-- Added the following to nameLink tag:
 ELEMENTS
   collectionName
   eventName
   materialName
   objectName
   titleName
   objectIdentifier

 ATTRIBUTES
```

```
   authority
   role
-->

<!ELEMENT nameLink           (%phrase.seq;)                        >
<!ATTLIST nameLink           %a.global;
                             %a.personPart;
          authority          (Non-Authoritative | Local-to-Server | AAT
| AAT.Date |
                             ACRL.RBMS-Binding | ACRL.RBMS-Genre |
ACRL.RBMS-Paper | ACRL.RBMS-Printing
                             | ACRL.RBMS-Type | Base-Merimee | BGN |
British-Archaelogical |
                             Canadiana | CDWA | Dictionarium-
Museologicu | Garnier | Geosaurus
                             | Glass | ICOM.Costume | Iconclass |
Jewish-Art | ISO.Language |
                             ISO.Documentation | ISO.Iconic | ISO.AV |
ISO.Date-Time | LC.Descriptive-Graphic
                             | LC.Name | LC.Thesaurus-Graphic | LCSH |
Moving-Image-Materials |
                             Nomenclature | Reynies | TGN | Tozzer |
Ulan | USMarc | Villard
                             | Yale.British-Artists)  #IMPLIED
          role               CDATA                #IMPLIED
          TEIform            CDATA                'nameLink'       >


<!-- Added the following to addName tag:
 ELEMENTS
   collectionName
   eventName
   materialName
   objectName
   titleName
   objectIdentifier

 ATTRIBUTES
   authority
   role
-->

<!ELEMENT addName            (%phrase.seq;)                        >
<!ATTLIST addName            %a.global;
                             %a.personPart;
          authority          (Non-Authoritative | Local-to-Server | AAT
| AAT.Date |
                             ACRL.RBMS-Binding | ACRL.RBMS-Genre |
ACRL.RBMS-Paper | ACRL.RBMS-Printing
                             | ACRL.RBMS-Type | Base-Merimee | BGN |
British-Archaelogical |
                             Canadiana | CDWA | Dictionarium-
Museologicu | Garnier | Geosaurus
                             | Glass | ICOM.Costume | Iconclass |
Jewish-Art | ISO.Language |
                             ISO.Documentation | ISO.Iconic | ISO.AV |
ISO.Date-Time | LC.Descriptive-Graphic
```

```
                              | LC.Name | LC.Thesaurus-Graphic | LCSH |
Moving-Image-Materials |
                              Nomenclature | Reynies | TGN | Tozzer |
Ulan | USMarc | Villard
                              | Yale.British-Artists)  #IMPLIED
          role                CDATA                 #IMPLIED
          TEIform             CDATA                 'addName'       >


<!-- Added the following to placeName tag:
 ELEMENTS
   collectionName
   eventName
   materialName
   objectName
   titleName
   objectIdentifier

 ATTRIBUTES
   authority
   role
-->

<!ELEMENT placeName           ((#PCDATA | %m.placePart; |
                              %m.phrase; )+)                    >
<!ATTLIST placeName           %a.global;
          type                CDATA                 #IMPLIED
          full                (yes | abb | init)    yes
                              %a.names;
          authority           (Non-Authoritative | Local-to-Server | AAT
| AAT.Date |
                              ACRL.RBMS-Binding | ACRL.RBMS-Genre |
ACRL.RBMS-Paper | ACRL.RBMS-Printing
                              | ACRL.RBMS-Type | Base-Merimee | BGN |
British-Archaelogical |
                              Canadiana | CDWA | Dictionarium-
Museologicu | Garnier | Geosaurus
                              | Glass | ICOM.Costume | Iconclass |
Jewish-Art | ISO.Language |
                              ISO.Documentation | ISO.Iconic | ISO.AV |
ISO.Date-Time | LC.Descriptive-Graphic
                              | LC.Name | LC.Thesaurus-Graphic | LCSH |
Moving-Image-Materials |
                              Nomenclature | Reynies | TGN | Tozzer |
Ulan | USMarc | Villard
                              | Yale.British-Artists)  #IMPLIED
          role                CDATA                 #IMPLIED
          TEIform             CDATA                 'placeName'    >


<!-- Added the following to settlement tag:
 ELEMENTS
   collectionName
   eventName
   materialName
   objectName
   titleName
```

```
   objectIdentifier

 ATTRIBUTES
   authority
   role
-->

<!ELEMENT settlement          (%phrase.seq;)                    >
<!ATTLIST settlement          %a.global;
                              %a.placePart;
          authority           (Non-Authoritative | Local-to-Server | AAT
| AAT.Date |
                              ACRL.RBMS-Binding | ACRL.RBMS-Genre |
ACRL.RBMS-Paper | ACRL.RBMS-Printing
                              | ACRL.RBMS-Type | Base-Merimee | BGN |
British-Archaelogical |
                              Canadiana | CDWA | Dictionarium-
Museologicu | Garnier | Geosaurus
                              | Glass | ICOM.Costume | Iconclass |
Jewish-Art | ISO.Language |
                              ISO.Documentation | ISO.Iconic | ISO.AV |
ISO.Date-Time | LC.Descriptive-Graphic
                              | LC.Name | LC.Thesaurus-Graphic | LCSH |
Moving-Image-Materials |
                              Nomenclature | Reynies | TGN | Tozzer |
Ulan | USMarc | Villard
                              | Yale.British-Artists)  #IMPLIED
          role                CDATA                #IMPLIED
          TEIform             CDATA                'settlement'  >


<!-- Added the following to region tag:
 ELEMENTS
   collectionName
   eventName
   materialName
   objectName
   titleName
   objectIdentifier

 ATTRIBUTES
   authority
   role
-->

<!ELEMENT region              (%paraContent)                     >
<!ATTLIST region              %a.global;
                              %a.placePart;
          authority           (Non-Authoritative | Local-to-Server | AAT
| AAT.Date |
                              ACRL.RBMS-Binding | ACRL.RBMS-Genre |
ACRL.RBMS-Paper | ACRL.RBMS-Printing
                              | ACRL.RBMS-Type | Base-Merimee | BGN |
British-Archaelogical |
                              Canadiana | CDWA | Dictionarium-
Museologicu | Garnier | Geosaurus
```

```
                              | Glass | ICOM.Costume | Iconclass |
Jewish-Art | ISO.Language |
                              ISO.Documentation | ISO.Iconic | ISO.AV |
ISO.Date-Time | LC.Descriptive-Graphic
                              | LC.Name | LC.Thesaurus-Graphic | LCSH |
Moving-Image-Materials |
                              Nomenclature | Reynies | TGN | Tozzer |
Ulan | USMarc | Villard
                              | Yale.British-Artists)  #IMPLIED
          role                CDATA                #IMPLIED
          TEIform             CDATA                'region'       >


<!-- Added the following to country tag:
 ELEMENTS
   collectionName
   eventName
   materialName
   objectName
   titleName
   objectIdentifier

 ATTRIBUTES
   authority
   role
-->

<!ELEMENT country             (%paraContent)                        >
<!ATTLIST country             %a.global;
                              %a.placePart;
          authority           (Non-Authoritative | Local-to-Server | AAT
| AAT.Date |
                              ACRL.RBMS-Binding | ACRL.RBMS-Genre |
ACRL.RBMS-Paper | ACRL.RBMS-Printing
                              | ACRL.RBMS-Type | Base-Merimee | BGN |
British-Archaelogical |
                              Canadiana | CDWA | Dictionarium-
Museologicu | Garnier | Geosaurus
                              | Glass | ICOM.Costume | Iconclass |
Jewish-Art | ISO.Language |
                              ISO.Documentation | ISO.Iconic | ISO.AV |
ISO.Date-Time | LC.Descriptive-Graphic
                              | LC.Name | LC.Thesaurus-Graphic | LCSH |
Moving-Image-Materials |
                              Nomenclature | Reynies | TGN | Tozzer |
Ulan | USMarc | Villard
                              | Yale.British-Artists)  #IMPLIED
          role                CDATA                #IMPLIED
          TEIform             CDATA                'country'    >


<!-- Added the following to bloc tag:
 ELEMENTS
   collectionName
   eventName
   materialName
   objectName
```

```
   titleName
   objectIdentifier

 ATTRIBUTES
   authority
   role
-->

<!ELEMENT bloc                (%phrase.seq)                        >
<!ATTLIST bloc                %a.global;
                              %a.placePart;
          authority           (Non-Authoritative | Local-to-Server | AAT
| AAT.Date |
                              ACRL.RBMS-Binding | ACRL.RBMS-Genre |
ACRL.RBMS-Paper | ACRL.RBMS-Printing
                              | ACRL.RBMS-Type | Base-Merimee | BGN |
British-Archaelogical |
                              Canadiana | CDWA | Dictionarium-
Museologicu | Garnier | Geosaurus
                              | Glass | ICOM.Costume | Iconclass |
Jewish-Art | ISO.Language |
                              ISO.Documentation | ISO.Iconic | ISO.AV |
ISO.Date-Time | LC.Descriptive-Graphic
                              | LC.Name | LC.Thesaurus-Graphic | LCSH |
Moving-Image-Materials |
                              Nomenclature | Reynies | TGN | Tozzer |
Ulan | USMarc | Villard
                              | Yale.British-Artists)  #IMPLIED
          role                CDATA                #IMPLIED
          TEIform             CDATA                'bloc'        >


<!-- Added the following to geogName tag:
 ELEMENTS
   NONE

 ATTRIBUTES
   authority
   role
-->

<!ELEMENT geogName            (#PCDATA | geog | name)*          >
<!ATTLIST geogName            %a.global;
                              %a.placePart;
          authority           (Non-Authoritative | Local-to-Server | AAT
| AAT.Date |
                              ACRL.RBMS-Binding | ACRL.RBMS-Genre |
ACRL.RBMS-Paper | ACRL.RBMS-Printing
                              | ACRL.RBMS-Type | Base-Merimee | BGN |
British-Archaelogical |
                              Canadiana | CDWA | Dictionarium-
Museologicu | Garnier | Geosaurus
                              | Glass | ICOM.Costume | Iconclass |
Jewish-Art | ISO.Language |
                              ISO.Documentation | ISO.Iconic | ISO.AV |
ISO.Date-Time | LC.Descriptive-Graphic
```

```
                              | LC.Name | LC.Thesaurus-Graphic | LCSH |
Moving-Image-Materials |
                              Nomenclature | Reynies | TGN | Tozzer |
Ulan | USMarc | Villard
                              | Yale.British-Artists)  #IMPLIED
          role                CDATA                 #IMPLIED
          TEIform             CDATA                 'geogName'    >


<!-- Added the following to orgName tag:
 ELEMENTS
   collectionName
   eventName
   materialName
   objectName
   titleName
   objectIdentifier

 ATTRIBUTES
   authority
   role
-->

<!ELEMENT orgName             (%phrase.seq)                        >
<!ATTLIST orgName             %a.global;
          type                CDATA                 #IMPLIED
          key                 CDATA                 #IMPLIED
          reg                 CDATA                 #IMPLIED
          authority           (Non-Authoritative | Local-to-Server | AAT
| AAT.Date |
                              ACRL.RBMS-Binding | ACRL.RBMS-Genre |
ACRL.RBMS-Paper | ACRL.RBMS-Printing
                              | ACRL.RBMS-Type | Base-Merimee | BGN |
British-Archaelogical |
                              Canadiana | CDWA | Dictionarium-
Museologicu | Garnier | Geosaurus
                              | Glass | ICOM.Costume | Iconclass |
Jewish-Art | ISO.Language |
                              ISO.Documentation | ISO.Iconic | ISO.AV |
ISO.Date-Time | LC.Descriptive-Graphic
                              | LC.Name | LC.Thesaurus-Graphic | LCSH |
Moving-Image-Materials |
                              Nomenclature | Reynies | TGN | Tozzer |
Ulan | USMarc | Villard
                              | Yale.British-Artists)  #IMPLIED
          role                CDATA                 #IMPLIED
          TEIform             CDATA                 'orgName'     >
```

11 Appendix E – Unnecessary Tags in the Tag Set

This appendix contains all of the tags in the DTD that will not be used in the transcription of the Robinson Reports and, therefore, were not mentioned in Section 4.3. These tags were included because they are a part of one of the TEI tag sets that make up the DTD. They were not removed from the set for two reasons: first, it is a very tedious process to remove tags from a DTD; second, any deviation from the standard is undesirable. These tags will be described in brief here, along with the reason they will not be used; however, no detailed explanation of their use will be provided with this report. Tags that are exclusively used in the TEI header, but are not being used or suggested for future use in this project, will not be documented here. For more information on these tags see *TEI Guidelines for Electronic Text Encoding and Interchange (P3)*.

11.1 Document Structure

The elements in this section pertain to the structure of the text. These tags delineate front and back matter, as well as divisions within the text itself.

- **<front>** belongs before the <body> element, but within <text>. It contains any materials that occur before the beginning of the actual text, such as title pages and prefaces. This tag is not necessary, as the Robinson Reports have no front matter.

- **<back>** belongs after the <body> element, but within <text>. It contains any materials that occur after the end of the text, such as appendices, glossaries, and indices. This tag is not necessary, as the Robinson Reports have no back matter.

- **<group>** contains several texts that belong together for some reason, such as a collection of short stories. The group as a whole can have front and back matter, as can each of the individual texts within the group; however, the individual texts do not have their own headers. Although the Robinson Reports

will eventually be but into a larger composite, it will be done using the TEI Corpus; therefore, the <group> tag is unnecessary.

- **<div0>...<div7>** are used if there are multiple types of subdivisions in a text. <div0> represents the largest subdivision, while <div7> represents the smallest. These tags will not be used because the Robinson Reports have no subdivisions.

- **<milestone>** marks any sort of sections in the text that are not otherwise marked. <milestone> is used to mark divisions of the text that are indicated in some form of reference system. As the divisions in the Robinson Reports are being marked otherwise, the <milestone> tag will not be necessary.

- **<cb>** signifies the beginning of a new column in a text. The Robinson Reports have no columns, so the <cb> tag is unnecessary.

- **<argument>** denotes a list of topics found at the beginning of a division of the text. The Robinson Reports have no such lists, so this tag will not be used.

- **<epigraph>** signifies a quotation found at the beginning of a division of the text. The Robinson Reports have no epigraphs, so this tag will not be used.

- **<trailer>** identifies a footer or other closing text found at the end of a division of the text. The Robinson Reports do not have closing text, so this tag will not be used.

- **<fw>** is used to tag a header or footer that runs throughout a document. Even reports that are multiple pages do not have any kind of running header or footer, so <fw> will not be used for the transcription of the Robinson Reports.

11.2 Bibliographic Citation

The elements included in this section are used in bibliographic citation, but are not used in the TEI header of the Robinson Reports. These elements are used to mark any form of bibliographic information, including fully structured lists of cited works and loosely structured bibliographic information given within the text of a document.

- **<cit>** contains a <quote>, described in Section 4.2.1.1, and a bibliographic citation, if that citation is supplied with the quote in the text. <cit> will not be used because the Robinson Reports do not supply bibliographic citations with quotations used in them.

- **<biblStruct>** denotes a structured bibliographic citation. This contains only bibliographic elements, which are presented in a specific order.

- **<biblFull>** is used for a fully structured bibliographic citation that contains all of the elements of a TEI file description.

- **<listBibl>** contains a list of bibliographic citations, which could be made up of <bibl>, <biblStruct>, and <biblFull> elements.

- **<analytic>**, **<monogr>**, and **<series>** are subelements of citation tags and contain information about different types of sources

- **<author>** is a sub-element of the citation tags that designates the name of the author of a work.

- **<editor>** is a sub-element of the citation tags that denotes the editor of a work.

- **<meeting>** is a sub-element of the citation tags that contains the name of the meeting or conference from where the item being cited was derived

- **<imprint>** is a sub-element of the citation tags, which contains information regarding publication of an item.

- **<pubPlace>** is a sub-element of <imprint> that describes the place of publication.

- **<publisher>** is a sub-element of <imprint> that holds the publisher's name.

- **<idno>** is a sub-element of the citation tags that designates an identifying number for the piece. This could be a standard number, such as an ISBN, or a contrived number.

- **<extent>** is a sub-element of the citation tags, which describes the size of electronic text.

- **<edition>** is a sub-element of the citation tags that is used for marking any information regarding the edition of the text.

11.3 Title Pages

The following elements are used for the documenting of title pages. These elements are generally used in the front matter of a document, but can be used in the back

matter as well. Because the Robinson Reports have no title pages, none of these tags will be used in their transcription.

- **<titlePage>** is the main element used to designate a title page. It can contain any of the subelements in this section, as well as general text elements. For instance, elements that describe publication information might be used on a title page.

- **<docTitle>** denotes the title of the document, as it is listed on the title page.

- **<titlePart>** is a sub-element of <docTitle> that signifies a subtitle, or other division of the title.

- **<byline>** identifies the name of the person primarily responsible for the document. This includes any description of what the person did, such as "written by."

- **<docAuthor>** is a sub-element of <byline> that denotes specifically the author's name.

- **<imprimatur>** is used to tag any statement on the title page, which authorises the publication of the work.

- **<docEdition>** identifies the edition of the work, as it is listed on the title page.

- **<docImprint>** denotes the publication information of a work, as it listed on the title page. The bibliographic citation elements in Section 4.2.4.2 can be used as subelements of <docImprint> to more explicitly tag the publication information.

- **<docDate>** signifies the date of publication of the work, as given on the title page.

11.4 Glossaries and Indices

The elements described in this section are used for marking up and generating glossaries and indices. These are usually found in the back matter of a work, but can be designated as front matter as well.

- **<gloss>** is used to mark a definition of another word; <gloss> can be linked to a <mentioned> or <term> tag to indicate the word that the definition refers to. Robinson did not regularly supply definitions of words, so <gloss> is unnecessary.

- **<index>** is used for marking words in the text that will be added to an auto-generated index. No plans exist to generate an index for the Robinson Reports, so this tag will not be used.

- **<divGen>** indicates where a division in generated text should appear. This can be used to structure auto-generated indices. Because no plans exist for auto-generated text, the <divGen> tag will not be used.

11.5 Editorial Changes to the Text

The first three tags described here are alternatives to the tags described in Section 4.2.1.4. They are used to make editorial changes to the text, but are not being used in favour of <orig>, <abbr>, and <sic>, which accomplish the same goal in a slightly different manner. The last two tags are for corrections that were made on the original document, and are alternatives to tags described in Section 4.3.1.3.

- **<reg>** indicates text that has been changed by the transcriber to modern spelling, or otherwise regularised. <reg> will not be used, in favour of the <orig> tag.

- **<expan>** is used where the transcriber has replaced an abbreviation with its expansion. <expan> will not be used, in favour of the <abbr> tag.

- **<corr>** is used where the transcriber has replaced a known mistake with corrected text. <corr> will not be used, in favour of the <sic> tag.

- **<delSpan>** is similar to <del>, but is used for longer passages that have been deleted. <delSpan> will not be used in favour of <del>.

- **<addSpan>** is similar to <add>, but is used for longer passages that have been added in. For instance, a footnote written added by the author to expound on a point would be marked with <addSpan>. <addSpan> will not be used in favour of <add>.

11.6 Fiction, Verse and Drama

The tags described in this section are specifically used in fiction, verse, and drama. Since the Robinson Reports contain none of these things, all of these tags will be unnecessary for their transcription.

- **<q>** is used to mark quotations that occur within the text, such as dialogue between characters.
- **<l>** signifies a single line of verse.
- **<lg>** contains a group of verse lines, such as a stanza.
- **<sp>** denotes a speech in either a play or in prose.
- **<speaker>** gives the name of a speaker in a play.
- **<stage>** indicates a stage direction.

11.7 Names, Addresses and Formulae

The elements described here are for tagging names, addresses and formulae. These elements are either irrelevant to the Robinson Reports or have been replaced by several more specific tags.

- **<rs>** stands for referring string, which can refer to any noun. <rs> will not be necessary because the specific tags – <persName>, <placeName>, <orgName>, <objectName>, and so forth – will be used in its place.
- **<street>**, **<postCode>**, and **<postBox>** are subelements of <address> that can be used as alternates to <adrLine>. When used in conjunction with <name>, which signifies cities, countries, and organisations, they break up an address into individual pieces. Since <address> and <name> will not be used, these elements will not be used.
- **<formula>** contains a mathematical formula. Mathematical formulae to not appear in the Robinson Reports, so this tag will not be used.

11.8 Dates and Times

This section describes the elements used for tagging dates and times in various formats, as well as elements that describe the order in which events happened. Times mentioned in the Robinson Reports are impertinent to the transcriptions and the dates will be designated with the basic <date> tag; therefore, the time elements and the remaining date elements are unnecessary.

- **<dateline>** contains information describing the place, date, and time of the writing of a text. The <dateline> tag is used as part of the <opener> or <closer> in the body of the text.

- **<dateStruct>** represents a date and allows for the individual parts to be broken down.

- **<day>**, **<week>**, **<month>**, and **<year>** are subelements of <dateStruct> that separate the individual parts of the date.

- **<time>** is used to mark a time of day, written in any format.

- **<timeRange>** contains two times or a phrase indicating a range of time.

- **<timeStruct>** represents a date and allows for the individual parts to be broken down.

- **<second>**, **<minute>**, and **<hour>** are subelements of <timeStruct> that are used to tag the individual parts of the time.

- **<occasion>** can be a sub-element of either <timeStruct> or <dateStruct>. It signifies an event in time, such as Christmas.

- **<timeline>** is used to specify the order of events; it is usually used in transcriptions of speech to describe when words were said, and if they were said at the same time.

- **<when>** specifies a point in time. <when> can be used as a sub-element of <timeline> or on its own.

11.9 Tables and Lists

The tags listed here are used in the specification of tables and lists. A list in TEI contains a single listing of information, while a table can have any number of rows or columns. Both can have a heading specified within the <table> tag, using the <head> tag described in Section 4.3.2.2. Tables will be unnecessary in the transcriptions of the Robinson Reports; due to the complex nature of lists, lists will be used instead. The first three tags in this section deal with the creation of tables. The last two are tags used in creating lists that are not necessary for encoding the Robinson Reports. For a description of the list tags that will be used, see Section 4.3.1.1.1.

- **<table>** is the main element used in the creation of tables. It contains all of the heading, row, and cell elements.
- **<row>** is a sub-element of <table> that denotes the start of a new row within the table.
- **<cell>** is a sub-element of <table> and <row>, which signifies one cell within a table. The data held in the cell of a table would be contained within this element.
- **<headLabel>** is a sub-element of <list>, which indicates the heading of the label column in a list. This can be used in place of <head> when a list has a label column.
- **<headItem>** is a sub-element of <list> that contains the heading for item column if <headLabel> is used.

11.10 Segmentation

Links and other pointers are made by referring to the "id" attribute of another element; to facilitate linking, tags are provided to describe otherwise unmarked areas of the text. This segmentation of the text has many levels of detail, including a generic <seg> element for unclassified segments. The <seg> element is discussed in Section

4.3.3.5. None of the tags in this section will be used in the Robinson Reports because the level of analysis being done on them is not extremely detailed.

- **<s>** specifies a sentence.
- **<cl>** denotes a clause.
- **<w>** signifies a word, and allows for the inclusion of its lemma.
- **<m>** is used to tag a morpheme; <m> allows for the specification of the base form.
- **<c>** indicates a character.

11.11 Linking

The tags in this section are used in different methods of pointing to or referencing specific sections of the text. These methods will not be used in the Robinson Reports; however, other methods that may be used are described in Section 4.3.3.5.

- **<ptr>** supplies a pointer to another part of the text.
- **<ref>** denotes a reference to another part of the text.
- **<xref>** denotes a reference to an external text.
- **<join>** is used to join two parts of a broken segment, so that they can be treated as one. <join> effectively creates a link between the two segments.

11.12 Interpretation

The tags in this section are designed to allow the transcriber include brief interpretations of the text. These tags are not for long explanations, but rather they are for a descriptive phrase that describes the role of a particular span of text. For instance, the first few sentences in a speech could be designated as the introduction.

- **** associates an interpretive comment with a particular span of text.

- **<spanGrp>** is used to group a series of tags together. <spanGrp> is useful because it allows attributes to be specified for the entire group; hence, they do not have to be entered into each element.

- **<interp>** provides an interpretive comment that is linked to a particular span of text.

- **<interpGrp>** is used to group a series of <interp> tags together. Like <spanGrp>, <linkGrp> allows attributes to be specified for the entire group.

- **<alt>** allow the transcriber to enter multiple possibilities of what a span of text should say. This is particularly useful when transcribing spoken text and it is unclear what was said.

- **<altGrp>** is used to group a series of <alt> tags together. Like <spanGrp>, <altGrp> allows attributes to be specified for the entire group.

12 Appendix F – User Manual for the Transcriptions

Transcription Manual

written by

Chris Holt
Micah Kiffer
Keith Peterson

Completed as a part of an Interactive Qualifying Project for Worcester Polytechnic Institute, in conjunction with the National Art Library at the Victoria and Albert Museum in London

Table of Contents

1 GETTING STARTED 1

1.1 AN INTRODUCTION TO TAGGING 1

1.2 TAG HIERARCHIES 3

1.3 TAG ATTRIBUTES 4

1.4 DOCUMENT TEMPLATES 6

2 BASIC FEATURES OF WORDPERFECT 9 XML EDITOR 9

2.1 MANAGING THE FILES 9

2.1.1 Creating a New Document Using a Template 9

2.1.2 Saving a Document 11

2.1.3 Opening a Document 15

2.2 CHANGING SCREEN DISPLAYS 16

2.3 USING THE XML TREE WINDOW 17

2.3.1 Manipulating the View of the XML Tree 18

2.3.1.1 The Tag Hierarchy 18

2.3.1.2 The Tag Attribute 19

2.3.2 Using the XML Tree Editor to Find an Error in the Document 19

2.4 ADDING A TAG 20

2.4.1 Opening the Elements Dialog Box 21

2.4.2 Adding the Tag 22

2.5 EDITING AN ATTRIBUTE OF A TAG 23

2.5.1 Editing an Attribute using the XML Tree 23

2.5.2 Editing an Attribute Using the Menu Bar 24

2.6 INSERTING A FOOTNOTE 25

2.7 UNDOING THE LAST CHANGE 27

2.8 INSERTING A SYMBOL 28

3 TEMPLATES 31

3.1 COMMON CHARACTERISTICS 31

3.1.1 Prompts 31

3.1.2 Common Template Structure 33

3.1.2.1 TeiHeader Element 34

3.1.2.2 Back Section 35

3.1.2.3 Main Section 35

3.1.2.4 Minutes and Annotations 36

3.1.2.5 Comments 38

3.2 THE ADDENDUM TEMPLATE 39

3.2.1 Prompts 40

3.2.2 Main Section Structure 40

3.3 THE BASIC LETTER TEMPLATE 41

3.3.1 Prompts 41

3.3.2 Main Section Structure 42

3.4 THE FORM REPORT TEMPLATE 43

3.4.1 Prompts 43

3.4.2 Main Section Structure 44

3.5 THE LIST OF OBJECTS TEMPLATE 46

3.5.1 Prompts 46
3.5.2 Main Section Structure 47
3.6 THE MINUTE PAPER TEMPLATE 48
3.6.1 Prompts 49
3.6.2 Main Section Structure 49
3.7 THE TELEGRAPH TEMPLATE 51
3.7.1 Prompts 51
3.7.2 Main Section Structure 52
3.8 THE PROFREFPAPER TEMPLATE 53
3.8.1 Prompts 54
3.8.2 Main Section Structure 55
3.9 THE PROFREFPAPER(REC FROM) TEMPLATE 57
3.10 THE PRELIMINARY REPORTS TEMPLATE 58
3.10.1 Prompts 58
3.10.2 Main Section Structure 59
3.11 SAMPLE TRANSCRIPTIONS 60
4 A GUIDE TO THE TAGS 62
4.1 GLOBAL ATTRIBUTES 63
4.2 DOCUMENT STRUCTURE 67
4.3 NAMES 69
4.4 DATES 75
4.5 NUMBERS 77
4.6 DISTINCT TEXT 79
4.7 ILLEGIBLE OR DAMAGED TEXT 85
4.8 INTERNAL EDITORIAL REMARKS 89
4.9 EXTERNAL EDITORIAL REMARKS 92
4.10 DRAWINGS AND FIGURES 95
4.11 ANNOTATIONS, COMMENTS, AND INTERPRETATIONS 96
4.12 LISTS 98
4.13 AN INDEX TO THE TAGS 104
5 SHORTCUTS AND QUICKWORDS 106
6 APPENDIX A - REFERENCE NUMBERS FOR THE BOXES 107
7 APPENDIX B – SAMPLE LETTERS 108

1 Getting Started

This manual will explain how to use the WordPerfect 9 XML editor, in conjunction with a specially designed tag set and document templates, to transcribe the Robinson Reports. There are four main chapters to this document:

- Basic Features of WordPerfect 9 XML Editor
- Templates
- A Guide to the Tags
- Shortcuts and QuickWords

Before you can fully understand these chapters, and, hence, make use of the editing environment, you will need to understand several important concepts. These concepts include:

- Tags and tagging
- Empty tags
- Nesting
- Tag attributes
- Document templates

Each of these concepts will be explained in the remainder of this chapter.

1.1 An Introduction to Tagging

The system you will be using to transcribe the Robinson Reports does not employ formatting in the way conventional Word Processors do. Text will not appear in the editor like it will in the final display. Instead, it is marked with tags that describe the role of text, so that formatting can be applied to it later. For instance, you do not need to indent a paragraph, but rather tag it as a paragraph so that it can be interpreted and displayed with the indent later.

Tags, or elements as they are sometimes called, are denoted in the text by a start tag (opening tag) and an end tag (closing tag). The nature of the tag designates the role of the text between the two parts. To mark a paragraph, the opening tag, [1], must be placed at the beginning of the paragraph and the closing tag, , must be placed at the end. This can be seen in Figure 1.

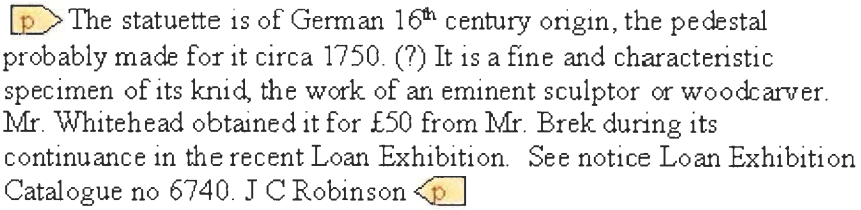

Figure 1 - Using a tag to mark a paragraph

There are more specific tags available to mark up the text of the Robinson Reports. These tags give meaning to the text by distinguishing the structure, content, style and tone, and physical appearance. Some of these tags will be used for display purposes, while others will facilitate searching and extracting relevant information from the transcriptions. Figure 2 shows the paragraph in Figure 1 after it has been fully marked up.

[1] In this manual, the notation is used synonymously with the p p notation used by WordPerfect. When discussing a tag name in the abstract, not as an example, only the opening tag, , will be used.

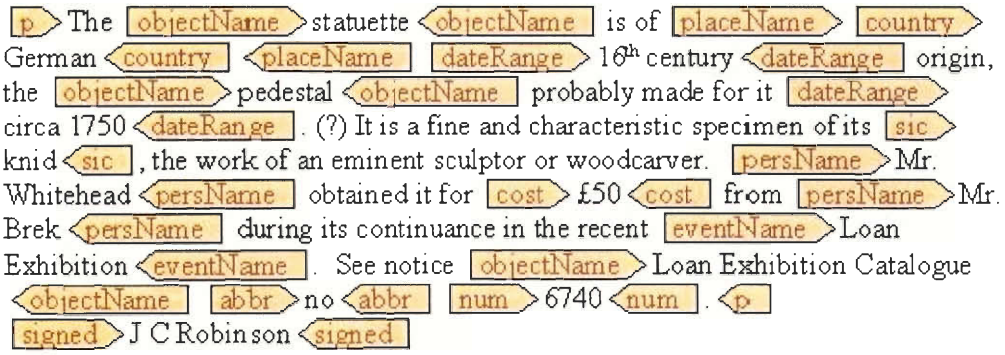

Figure 2 - A fully marked up paragraph

There are many more tags available, beyond those shown in Figure 2. Not all of the tags are pertinent to the Robinson Reports; however, each of the pertinent tags will be described in detail in chapter 4 *A Guide to the Tags*.

In review:

- **Tag** – *a mark inserted into the text to describe a particular aspect of the formatting, appearance, or content of the Robinson Reports; tags are often referred to as elements.*

1.2 Tag Hierarchies

Referring back to Figure 2, notice that the <country> tag on the first line is within the <placeName> tag. Some tags are allowed to appear within other tags to supply additional information about the text. This is called nesting.

Nesting is quite common, especially on a larger scale. Almost all tags appear within the paragraph tag, and paragraphs appear within the <body> tag (the <body> tag designates the body of the entire text). However, there are very specific rules that define which tags can be nested in each other tag. For instance, using the example from the beginning of this section, the <placeName> tag cannot be nested within the <country> tag. Likewise, the paragraph tag cannot be placed within the <placeName> tag, nor can the <body> tag be nested within a paragraph.

Similarly, some tags cannot have anything within them, not even text. These tags are referred to as empty tags. Empty tags generally provide some information that is relevant to a specific point in the text, such as a line break or a change in writing style. In WordPerfect, empty tags are displayed with the notation; however, this manual uses the same notation as regular tags.

All of the rules that define the tag hierarchy, from multiple nesting to empty tags, have been predefined and are enforced in the document templates with which you will be working; therefore, the editor will not allow you to insert tags where they don't belong.

In review:

- **Nesting** – *the process of placing one tag within another to more explicitly mark a part of the text; nesting must follow very specific rules, but these rules are enforced by the editor, which will alert you when they are broken.*

- **Empty Tags** – *tags in which the user cannot insert any text; empty tags usually mark changes that affect the remainder of the text.*

1.3 Tag Attributes

There are attributes associated with each tag that can be used to more explicitly describe the nature of the text within the tag. In order to fully understand all of the tags included for marking up the Robinson Reports, you will need to understand their attributes. Not every tag has the same attributes available to it; rather, each tag has attributes specifically designed to describe the type of text that it would be used to mark. For example, the <persName> tag (which is used to mark a person's name) has the type attribute available to it. This attribute can be used to denote the role of that person with regards to the Robinson Reports. This would be done in the following manner:

<persName type=artist>Donatello</persName>[2]

This signifies that Donatello was not only a person, but also an artist.

With attributes of the sort given in this example, where the value is text, acceptable values will generally be listed in chapter 4 *A Guide to the Tags*. If no values are listed, or you feel that none of the values given are appropriate, you can usually stray from these lists; however, it is very important that you do this with discretion. A large variety of attribute values will be of little use to a search engine. In addition, if you do create your own values for attributes, you must record your choices for the team that develops the search engine. Always remember, when creating attribute values, *be consistent* and *log your choices*.

It is also important to note that some tags have attributes of the same name. This does not necessarily mean that those attributes have the same function. An attribute common to two elements may have entirely different meanings for each. Because of this, it is important to refer to the documentation to learn how each attribute applies to each individual tag.

Although most tags have many attributes available to them, the majority of these attributes are not pertinent to the Robinson Reports; however, all of the attributes that are applicable will be documented with the tags to which they are associated in Chapter 4 *A Guide to the Tags*.

In review:

- **Attribute** – *descriptive information included within a tag to more explicitly describe the nature of the text; the attributes available to each tag are specified and enforced by the editor.*

[2] Although tag attributes are shown within the tag in this manual, WordPerfect does not use this notation. Attributes will not be displayed within the text of the transcriptions.

1.4 Document Templates

The document templates provide the fundamental structure for the transcriptions of the Robinson Reports. Each template corresponds to one of the document types shown in Table 1. They define the basic sections and elements that will appear in a transcription of a specific document type. It is important to note that the templates do not define all the tags that could be used for marking up the reports. They do however provide those tags that describe the general layout of the report. You must insert any additional tags that could be used to mark up the content during or after transcribing the written text.

Document templates not only provide the fundamental structure of the reports but introduce automation features as well. For example, when a new document is created from a template you will be prompted for information, which will automatically be inserted into the correct locations in the transcription document (For more information on prompts see section 3.1.1 *Prompts*). Detailed descriptions of each of the templates are provided in chapter 3 *Templates*.

Table 1 - Document types of the Robinson Reports

Document Type	**Explanation**	**Template**
Letter	The letter type represents a standard letter, containing a date, a body, and a signature. Letters come in various sizes and lengths, but all follow the same basic format. Occasionally, however, they include variations. These could be in the form of a greeting, a numbered or lettered list within the body of the letter, or a summary of the letter's contents. These summaries were written on the letters in red by a third-party.	Basic Letter
List	Lists represent reports that are pure lists, not lists included within the text of a letter. Examples of lists in the Robinson Reports objects in the museum, objects purchased for the art library, and objects with registry numbers and prices.	List of Objects

Table 1, cont'd - Document types of the Robinson Reports		
Document Type	**Explanation**	**Template**
Minute paper	Minute paper is a pre-printed form, which was used to acknowledge and give directives regarding other reports. This form includes three columns, one for names, one for dates, and one for minutes. However, these columns were often ignored, and had all three parts written in the minutes column. Minute paper has several variations.	Minute Paper
Telegraph	Telegraphs represent the various types of telegraphs that were used. Although there are a few different types, sent from different companies, all the telegraphs contain the same basic content.	Telegraph
Printed materials And sketches	Printed materials and sketches include any materials included with the reports that are not some form of hand written correspondence, either to or within the museum. These include printed catalogues of sales, full-page sketches, postcards, and printed pictures. These items either cannot or do not need to be transcribed.	Addendum
Professional Reference paper	Professional reference paper is a form letter that is very common beyond MA/3/28. These forms are specifically designed for reporting on works of art. They have separate fields for the name of the object, the asking price, the object's date and country of origin, and the body of the report. The body includes the Art Referee's recommendation regarding the object's purchase. Later variations of this form also include a field for the seller's name.	ProfRefPaper
Preliminary reports for the information of the director	Much like Professional Reference Paper, these reports are structured forms designed specifically for reporting on art objects. This form was used only for a short time, appearing in MA/3/28.	Preliminary Reports

Table 1, cont'd - Document types of the Robinson Reports		
Document Type	**Explanation**	**Template**
Form letter	The form letter reports are pre-printed letters with a few blank spaces left for the recipient's name, the topic of interest, and a signature. These reports referred to particular objects that were being considered for purchase. These objects were listed below the signature of the writer.	Form Report
Miscellaneous	Miscellaneous reports would include those that are not covered by the other types. All of the Robinson Reports should be covered in some manner by one of the types previously listed; however, a completely free-form document is necessary for any that are not.	Addendum

In review:

- **Document Template** – *a basis for the transcription and encoding of a particular document type; templates provide the overall structure of the reports.*

2 Basic Features of WordPerfect 9 XML Editor

This chapter of the user manual explains all the basic features of the WordPerfect 9 XML editor that you will need to transcribe the Robinson Reports. This chapter outlines how to manage the documents, add a tag, edit an attribute of a tag, change screen displays, use the XML Tree Window, insert a footnote, and undo the last change in the document.

2.1 Managing the Files

This section of the user manual will explain all the features necessary to manage your documents. These include: how to create a new document, how to save a document, and how to open an existing document. When you open or create a new document, you may want to consider opening the **Elements** window. This window provides you with all the valid tags available within the section of the document in which you are currently working. To find out how to open this window, refer to section 2.4.1 *Opening the Elements Dialog Box.*

2.1.1 Creating a New Document Using a Template

Every time you start a new transcription you will need to create a new document. This can be accomplished using the following steps.

1. Click on **File** and then **New from Project** (See Figure 3).

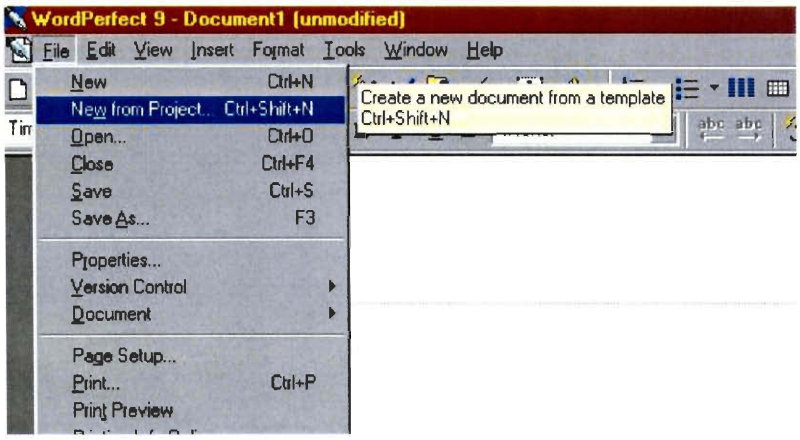

Figure 3 - Creating a new document

2. A dialog box will open, allowing you to choose a category (See Figure 4). Click on the top ▼ and choose **Transcriptions** from the drop-down menu.

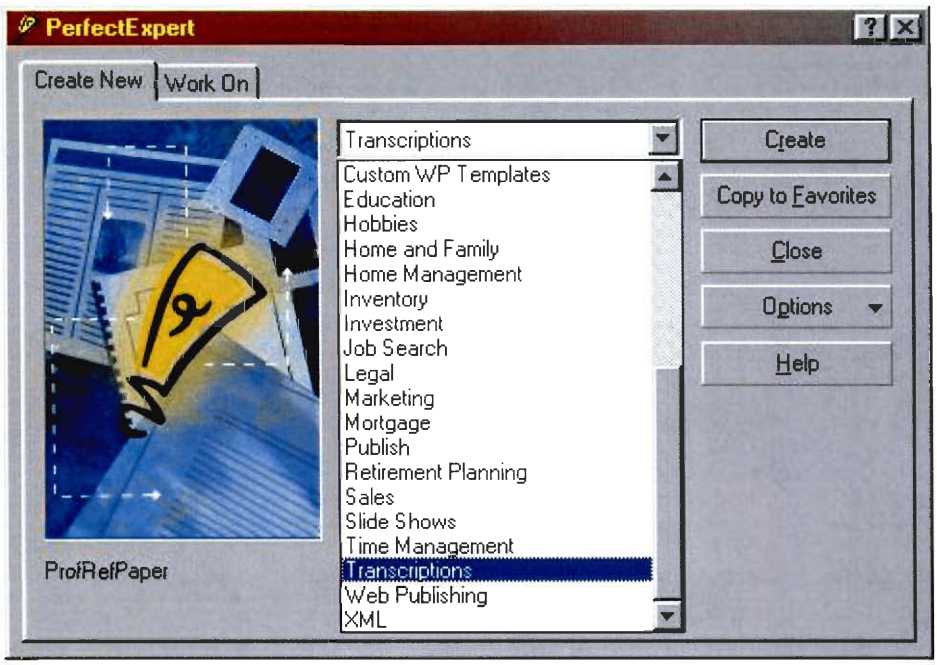

Figure 4 - Choose category

3. The available templates will appear in the list box shown in Figure 5. Click on the template that corresponds to the report you are transcribing.

4. Click **Create**

Figure 5 - Available Templates

2.1.2 Saving a Document

After you have finished writing or editing a transcription, you will want to save it. Use the following steps to complete this operation:

1. Click on **File** and then **Save** (See Figure 3), or hold down **Ctrl** and press **S**, or press in the *Property* toolbar.

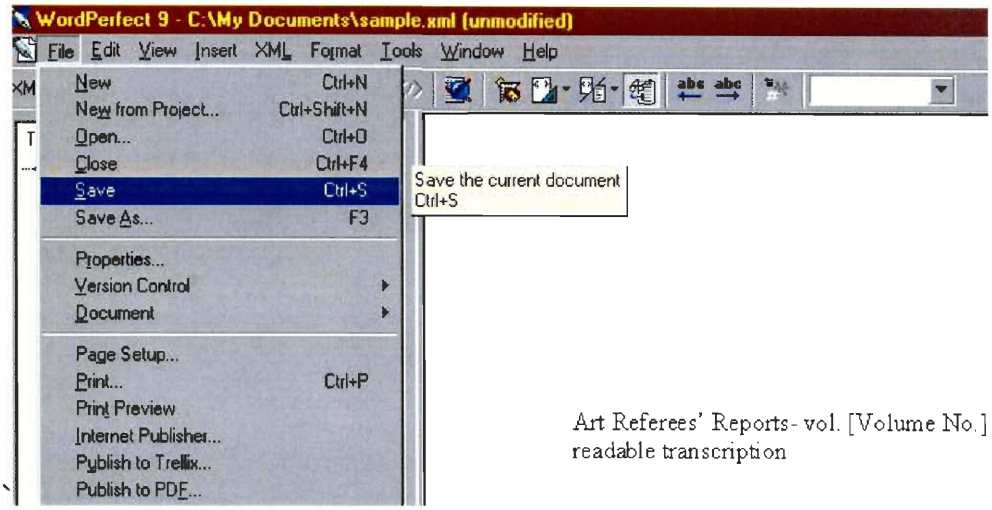

Figure 6 - Saving the document

***If you are saving the document for the first time complete step 2a. If not complete step 2b. After you have completed either of the steps go to step 3.

2a. A dialog box will appear asking you to name your transcription (See Figure 7). Before saving the document, make sure you are in the folder *C:\My Documents*. Use the following convention to name the transcriptions:

Naming Convention

[Box Number]-[Item number of the first page in the report].

e.g. MA-3-4-3

***If the item number is in the form #/#, use #-#

e.g. MA-3-46-97-1 should be used if the number on the report is 97/1

***Note: Make sure that **XML UTF-8** is selected under the *File Type* option.

***Note: Make sure that you use the correct box numbers. The correct #'s can be found in Appendix A.

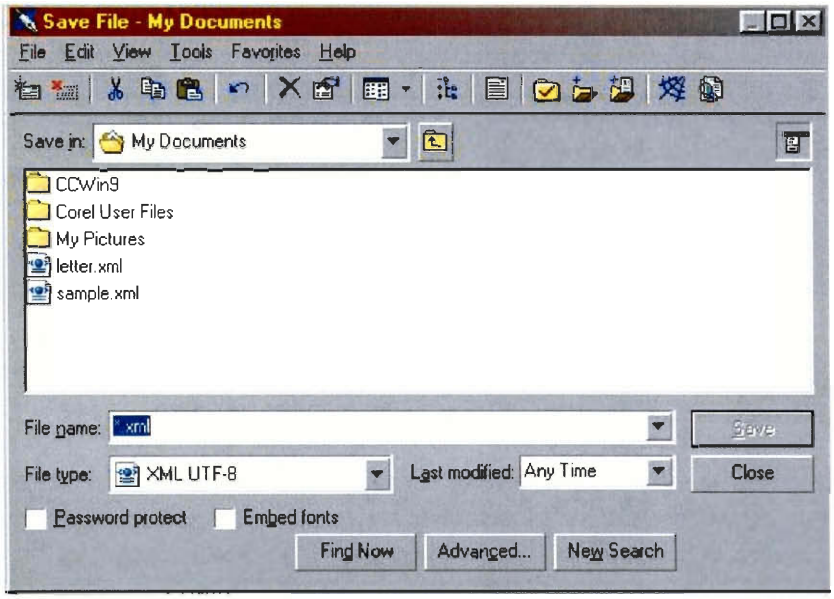

Figure 7 - Saving the transcription for the first time

2b. A window will pop up allowing you to choose what format you want to save the document in (See Figure 8). Make sure **XML UTF-8** is chosen and then click **OK**.

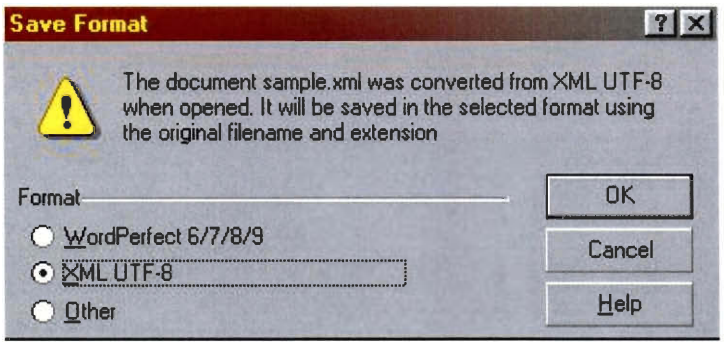

Figure 8 - Selection of save format

3. An additional dialog box will appear providing a few options on how to save the document (See Figure 9). Make sure that **Check For Errors Before Saving** is the **only** option selected and then click **OK**.

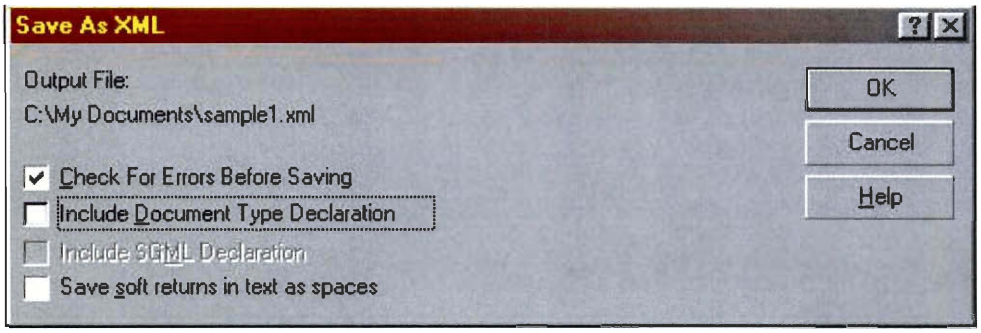

Figure 9 - Options for saving the document

***Note: If you are prompted with an error dialog box (Shown in Figure 10), complete steps 4, 5, and 6. If an error dialog appears up it will inform you of what is causing the error. Also you can view the place of the error using the XML tree. To get more information on how to use the XML tree to view errors see section 2.3.2. If an error box does not appear then only complete step 5.

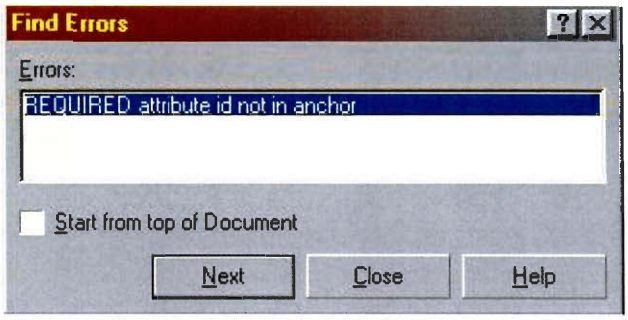

Figure 10 - Error dialog box

4. The cursor will be moved to the spot of the error. Fix the error then click **Next**. Repeat this step until you are prompted with a dialog box telling you that there are no more errors in file (See Figure 11). Click **Yes**. If you are prompted with a dialog box similar to Figure 10 then complete this step again. If not jump to step 5.

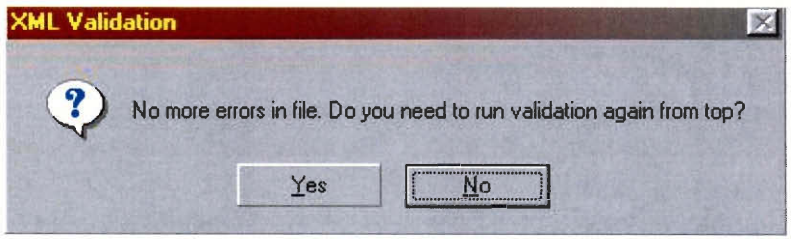

Figure 11 - XML validation window.

5. Click **No** in the *XML Validation* window (See Figure 11).

6. A window will prompt you telling you that errors were found in the document (See Figure 12). Assuming all errors were corrected, click **Yes**.

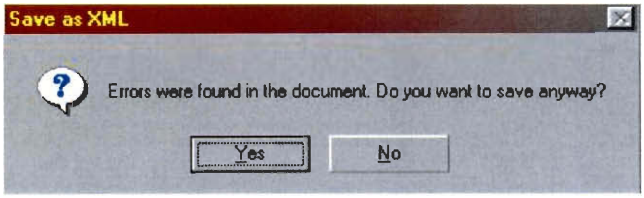

Figure 12 - Errors found prompt

2.1.3 Opening a Document

If at any time you need to go back to a transcription, whether it be to add comments or to finish a transcription you were in the middle of at the end of the day, you will need to open that document so that you can finish it. Complete the following steps to open a document.

1. Click on **File** and then click **Open** (See Figure 13), or hold down **Ctrl** and press **O**, or click on in the *Property* toolbar.

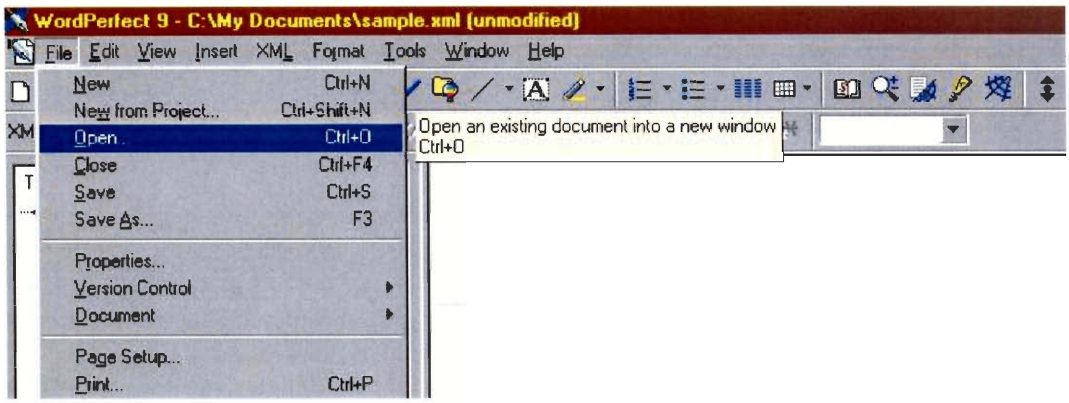

Figure 13 - Opening a document

2. Choose the document that you wish to open and click **Open** (See Figure 14).

***Note: make sure that you are in the **My Documents** folder. If you are not, change to the folder *C:\My Documents*.

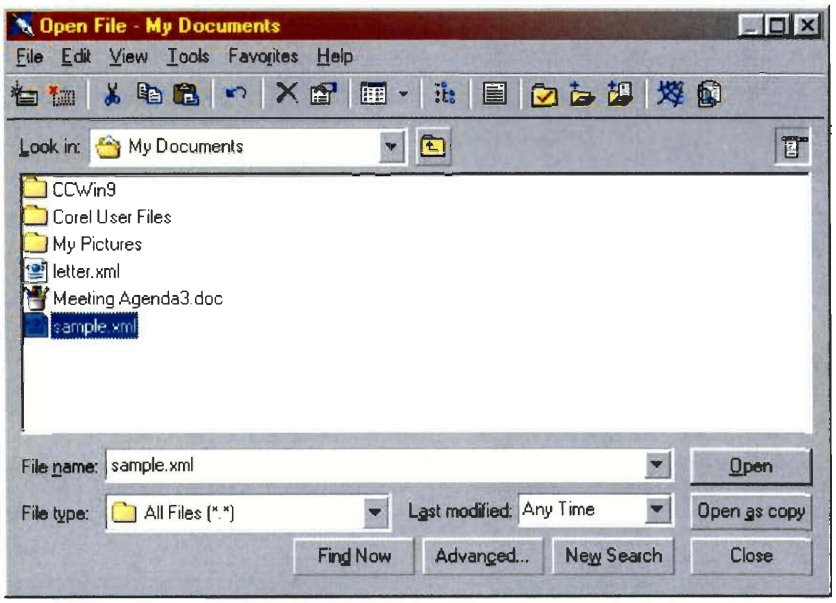

Figure 14 - Open file window

2.2 Changing Screen Displays

The WordPerfect 9 XML editor provides three different views of the document. The default view is called *Display Codes.* This view allows you to see all the tags in the document. Another view *Display as Markers,* removes all the tags from the document view but puts brackets in to signify the opening and closing of a tag. The final view available is the *Hide Codes* view. This view also hides all of the tags, but does not provide brackets. Use the following steps to switch between the three views.

1. Click on the button , which can be found in the *property bar* (See Figure 15)
2. Choose the view that you want.

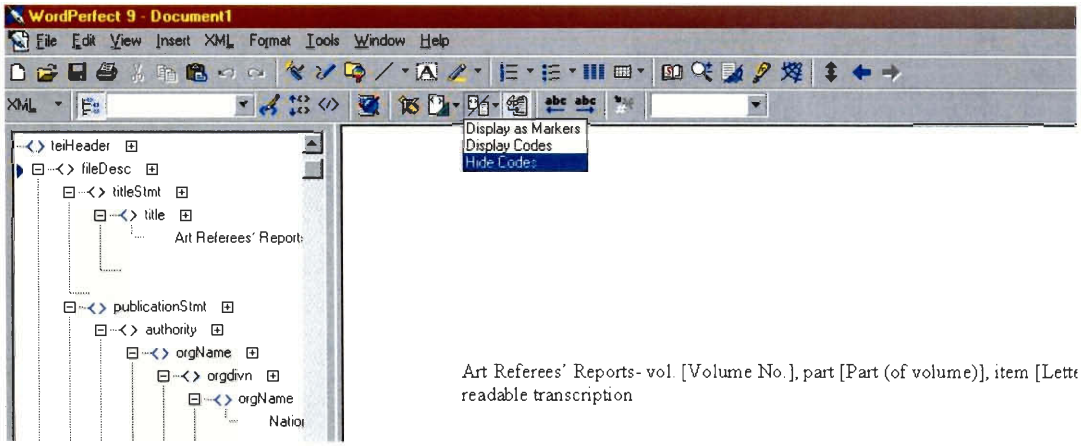

Figure 15 - Changing screen displays

2.3 Using the XML Tree Window

This section of the user manual will explain how to use the XML tree window. The XML tree window shows a hierarchical structure of the tags in the document. It allows you to edit attributes of tags and shows you where errors occur in the document. It also allows you to progress through the document easily by clicking on the tag you want to edit. The XML tree can be seen in Figure 16. Click on the button on the *Property* toolbar to open the XML tree.

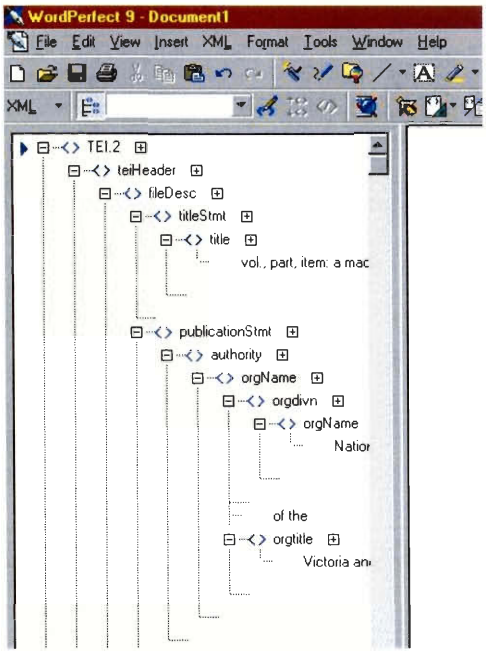

Figure 16 - XML Tree

2.3.1 Manipulating the View of the XML Tree

The XML Tree allows you to view the tag hierarchy as well as view and edit the attributes of tags. The editing of attributes using the XML tree will be explained in Section 2.5.1.

2.3.1.1 The Tag Hierarchy

Every tag can be expanded and reduced according to the user's preference (See Figure 17). When the tag is expanded, the XML tree will show all the tags nested within it. To view the tags click on the box to the left of the tag. If a "–" sign is in the box, the tag is expanded; if a "+" sign is in the box, then it is reduced.

Figure 17 - Tag in the XML tree

2.3.1.2 The Tag Attribute

To edit the attributes of a tag, click on the box to the right of the tag. If a "–" sign is in the box, the tag is expanded; if a "+" sign is in the box, it is reduced. There are two views of the attributes. One view only shows the attributes that have a value entered and the other shows all the attributes of the selected tag. If you want to change view you must complete the following steps:

1. Right-click on a tag in the XML tree.

2. A menu will open. Move the cursor to **Options**. (See Figure 18). If a checkmark is to the left of **All Attributes** then all the attributes are displayed for each tag, otherwise only the attributes with a value entered will be displayed.

3. Click on **All Attributes** to change the view.

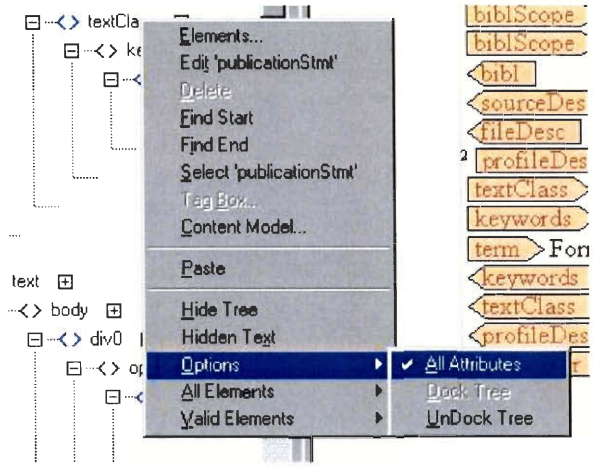

Figure 18 - Changing the attributes view in the XML tree

2.3.2 Using the XML Tree Editor to Find an Error in the Document

The XML Tree Editor conveniently advises you if an error is found in your document. If an error is found, the XML Tree will signify this by marking the tag that is causing the error with an error symbol. It also tells you the specific error if you move the cursor over the error symbol. An example of a tag that has an error can be seen in Figure

19. Sometimes the tag that is marked with the error symbol is not the tag causing the problem. For instance, if a <signed> tag is inserted before a paragraph (<p>) tag within the <body> tag, the <p> tag will be marked with an error because it appears after a <signed> tag. The real problem is that a <signed> tag must appear after all <p> tags in the <body> tag, not that a <p> tag is not allowed in a <body> tag. Use great care when fixing errors.

Figure 19 - Tag in the XML Tree with an error

2.4 Adding a Tag

This section of the user manual describes how to add a tag to a document. There are two ways to add a tag. You can either highlight the information that needs to be tagged or add a tag and then insert the information within the tag. This section includes descriptions of how to open the insert element window and how to use this window to insert a tag.

***Note: Once you become familiar with the names of tags and where they can legally be placed, you can use a shortcut for inserting a tag. With this method, you will have to type the tag name in a dialog box. To use this method complete the steps below:

1. Hold down **Ctrl** and press **E**.
2. The resulting dialog box will look like Figure 20. Type in the tag you wish to add and then click **Validate**.

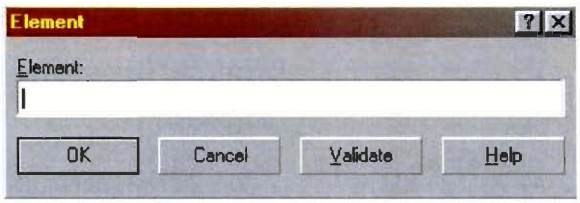

Figure 20 - Element dialog box

***Note: It is not necessary to validate the tag. If you are sure it is a real tag, you can skip this step.

***Note: If the tag is validated, it means it is a valid tag from the complete tag set. It does not mean that it is valid for the section of the document you are trying to in the tag. If an error symbol appears in the XML tree next to the tag, it is not valid for that section of the document.

3. A dialog box will appear telling you if the tag is valid or not. If it is valid then Figure 22 will appear. If it is not valid then Figure 21 will appear. Click **OK** on either of these boxes (See Figure 22).

Figure 21 - Invalid element dialog box

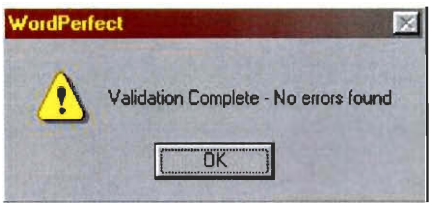

Figure 22 - Valid element dialog box

4. If the tag is valid, click **OK** to insert the tag (See Figure 20). If it is not valid go back to step 2.

2.4.1 Opening the Elements Dialog Box

This window can be easily opened and provides a valuable resource for you. It can display all of the available tags or only the tags that are valid in the section you are currently working on. This window will provide a useful resource while transcribing. It simplifies the process of adding new elements to the documents. Click **Insert** on the top toolbar and then click **Elements** to open the Elements dialog box (See Figure 23). If at any time you wish to close this dialog box click **Close**.

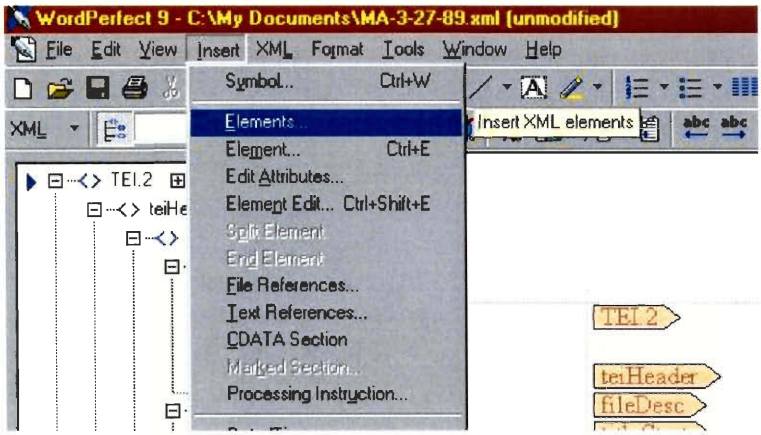

Figure 23 - Displaying element dialog

The resulting dialog box should look like Figure 24.

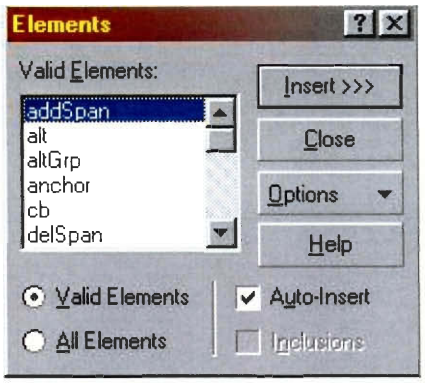

Figure 24 - Element Dialog

***Note: If you want to view all the tags click **All Elements** or if you only want to view the tags that are valid in the section you are working in click **Valid Elements**.

***Note: If this dialog box is open then the **Enter** key behaves differently when editing the document. You must press **Shift + Enter** to insert a carriage return and line feed.

2.4.2 Adding the Tag

If you wish to add a tag using the tag window you must complete the following steps:

1. Position cursor at the point in the document where you wish to insert the tag.

2. Progress through the list box until you find the tag name that you are looking for.

***Note: If you have chosen to view only valid elements the tag that you are looking for may not appear in the list if it is not valid to insert the tag at the point specified.

3. Click the tag name and the click **Insert**.

2.5 Editing an Attribute of a Tag

There are two ways to edit an attribute for a particular tag. The first requires the XML tree view to be open and the second uses the top tool bar. For more information on about working with the XML tree view, refer to section 2.3 *Using the XML Tree Window*. The editor can be set to automatically prompt you to edit attributes after inserting a new tag. There are three different options provided by the editor: **Always Prompt for Attributes**, **Never Prompt for Attributes,** and **Prompt When Attribute Required**. To select one of these options you must complete the following steps:

1. Open the elements dialog box as explained in section 2.4.1.
2. Click **Options** and choose the desired setting.

***Note: These settings do not work if you enter a tag by highlighting the text.

2.5.1 Editing an Attribute using the XML Tree

If you desire to add or correct an attribute using the XML Tree complete the following steps:

1. Move the cursor to the XML Tree on the left side of the screen, and locate the tag that you wish to edit.
2. If the attribute list is not visible, expand the attribute list of tag. For more information on how to expand the attribute list refer to section 2.3.1.2 *The Tag Attribute*

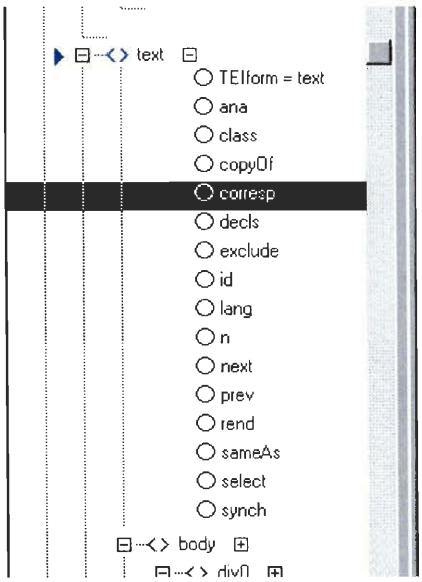

Figure 25 - XML tree tag expanded

3. Proceed through the list and double-click on the attribute you wish to edit (See Figure 25).

4. A dialog box will open up, allowing you to edit the value of the attribute (See Figure 26). Type in the value and click **OK**.

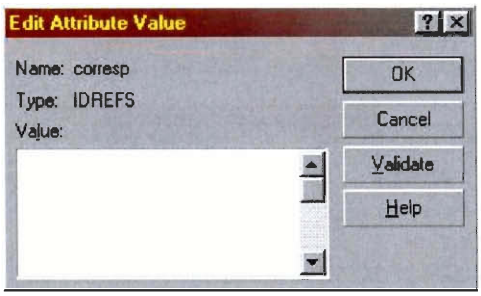

Figure 26 - Edit attribute dialog box

2.5.2 Editing an Attribute Using the Menu Bar

You can also use the menu bar to edit an attribute. This can be done using the following steps.

1. Position the cursor within the tag you wish to edit

2. Click **Insert** then click **Edit Attributes** (See Figure 27).

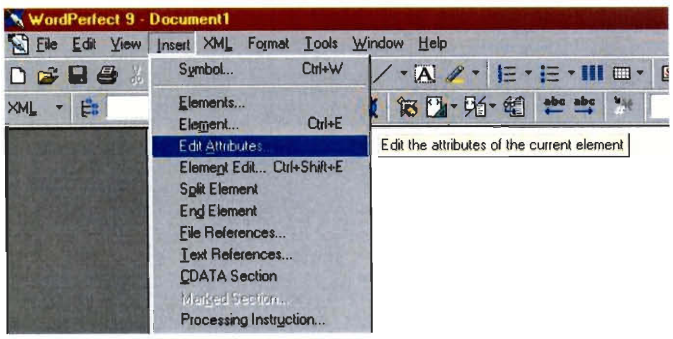

Figure 27 - Using the Top Menu to Edit Attributes

2.6 Inserting a Footnote

This section of the user manual describes how to add a footnote to transcriptions. There are two types of footnotes: one for comments provided by you and the other for minutes and annotations referring to an explicit point in the transcription. This section will explain how to add a footnote and how to differentiate the two types.

***Note: Footnotes can only be added to sections of the document where <note> tags can be included.

1. Click **Insert** in the menu toolbar and then click **Footnote/Endnote** (See Figure 28).

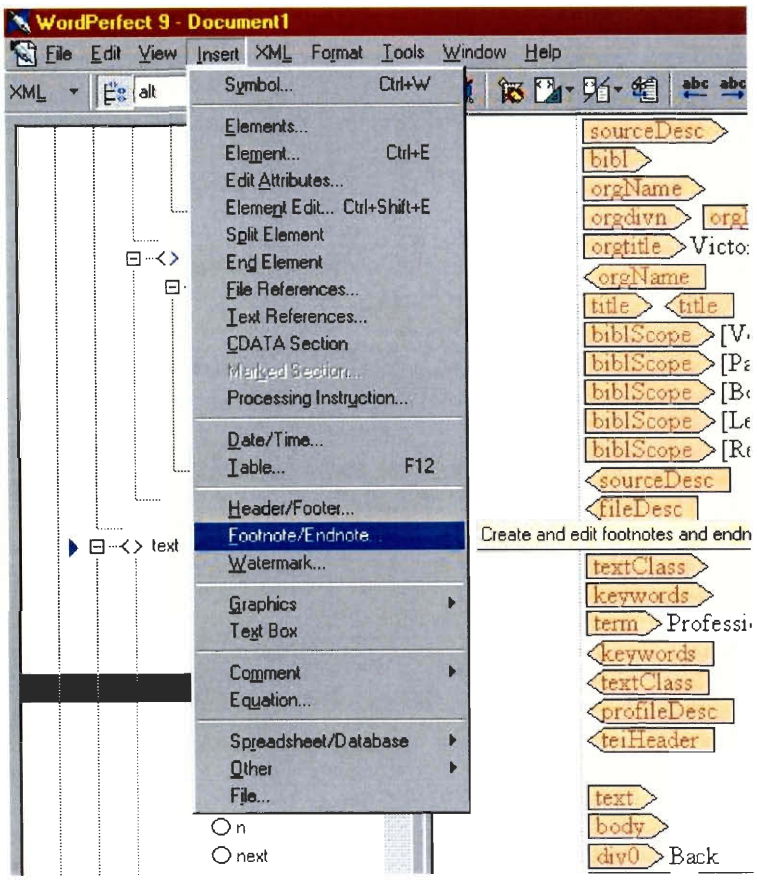

Figure 28 - Inserting a footnote

2. A dialog box will appear asking you if you want to insert a footnote or endnote. Make sure **Footnote** is selected and click **Create** (See Figure 29).

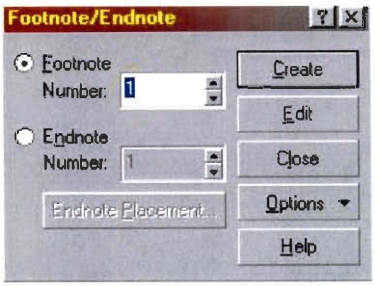

Figure 29 - Footnote/Endnote dialog box

3. The cursor will be brought to the bottom of the screen. Immediately insert a <note> tag.

4. Edit the attributes of the <note> tag.

***If the comments are provided by you, use the method below.

1. Edit the value of the <note> tag's attribute type; enter the text *comment.*
2. Edit the value of the <note> tag's attribute resp; enter the text *transcr.*
3. Edit the value of the <note> tag's attribute place; enter the text *footnote.*
4. Edit the value of the <note> tag's attribute anchored with *no.*

***If the comments are provided in the actual reports use the method below

1. Edit the value of the <note> tag's attribute type; enter the text *minute.*
2. Edit the value of the <note> tag's attribute resp; enter the name of the person who wrote the minute. If the name is unknown then type *unknown* in the resp attribute.
3. Complete steps 3 and 4 from above.

2.7 Undoing the Last Change

If you accidentally delete a tag or perform an undesirable operation while you are transcribing the reports you can easily undo that operation. There are two ways that you can undo an operation. You can use a shortcut or the menu bar. Using the shortcut is easy, simply hold down **Ctrl** and press **Z**. The menu bar is a little more time consuming but can be used until the shortcut is learned. Simply click **Edit** and then click **Undo** (See Figure 30).

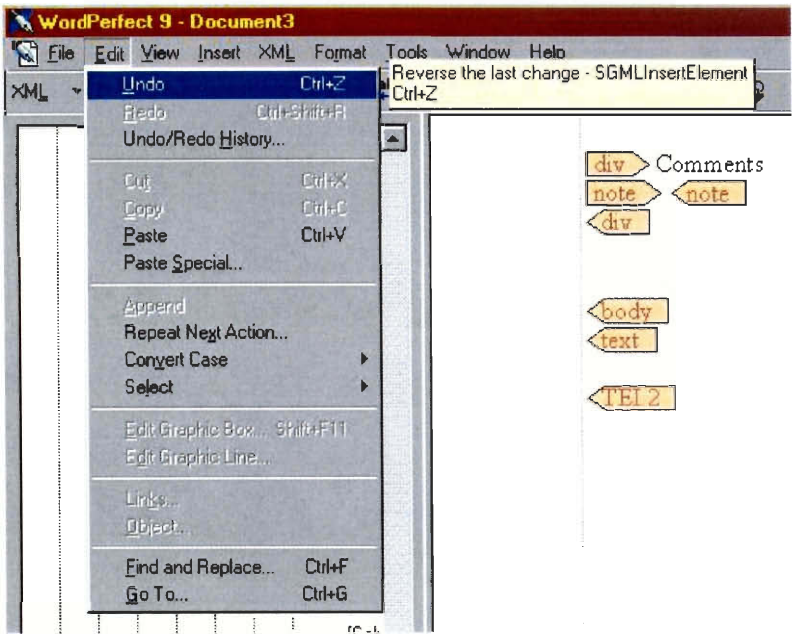

Figure 30 - Undoing the last change using the menu bar

2.8 Inserting a Symbol

You may need to insert a special character into one of the transcriptions. Special symbols include accented characters, typographic symbols, mathematical symbols, and other characters not appearing on the keyboard. To insert a special symbol, use the following steps:

1. Place the cursor where you wish to insert the symbol.
2. Click on **Insert** in the menu toolbar and then click **Symbol** (See Figure 31).

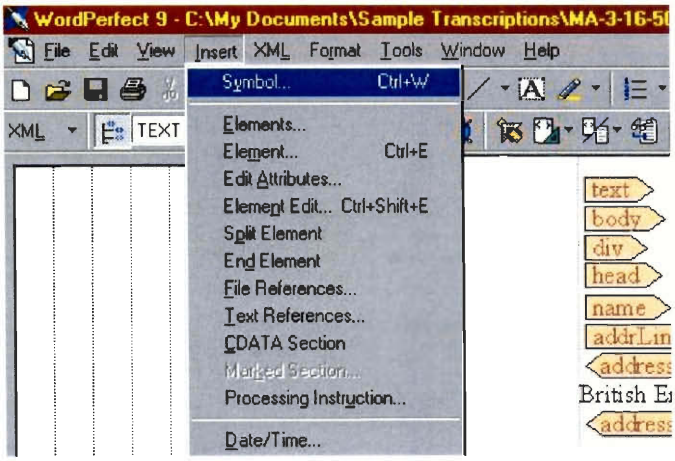

Figure 31 - Inserting a symbol

3. The symbols dialog box will appear (See Figure 32)

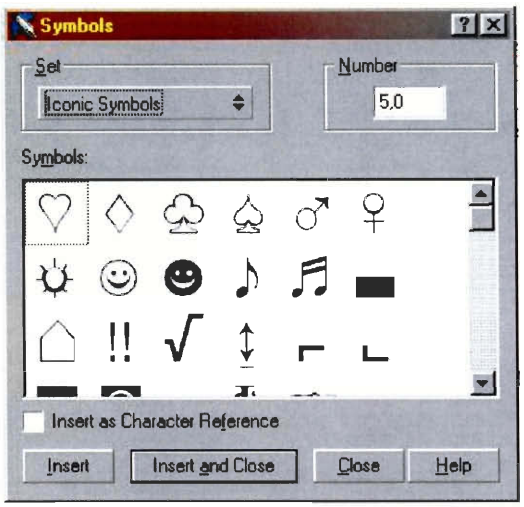

Figure 32 - The symbols dialog box

4. Click on the button below **Set** to choose a character set. Most of the symbols you will need to use are located in the *Multinational* and *Typographic Symbols* sets.

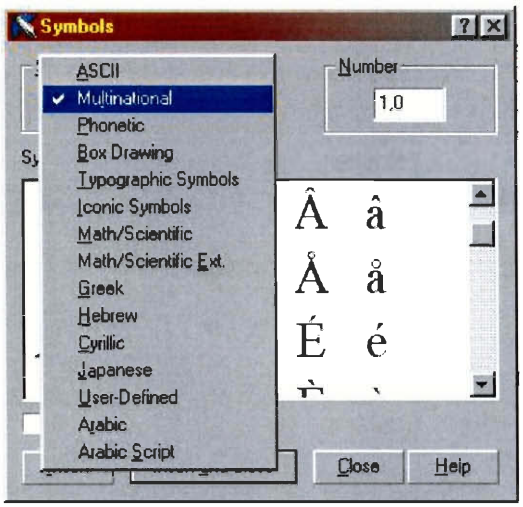

Figure 33 - Choosing a character set

5. Select the character you wish to insert and click **Insert and Close**. This will insert the symbol and close the dialog box.

3 Templates

Each of the templates provides a structure to follow for completing a transcription based on a certain document type. This structure does not define all of the tags that appear in a document; they only provide a starting point. Marking up the transcription text further will make the documents more useful to a system using them to extract relevant information. This chapter discusses in detail each of the templates created for the different document types. The following information will be provided for each template:

- Its corresponding document type
- Examples of reports that the template is used for
- A description of each of the prompts
- A description of its structure

***Note: When describing the structure of the templates all elements that need information to be added manually (not by prompts) will be in **bold**.

3.1 Common Characteristics

The templates were all designed in a similar fashion. You will find that each of the templates contains prompts and follows the same basic structure.

3.1.1 Prompts

When you create a new document from a template you will see a dialog box like the one show in Figure 34. This dialog box contains prompts, which are used to help

automate the entering of information that is required for or common to a particular document type.

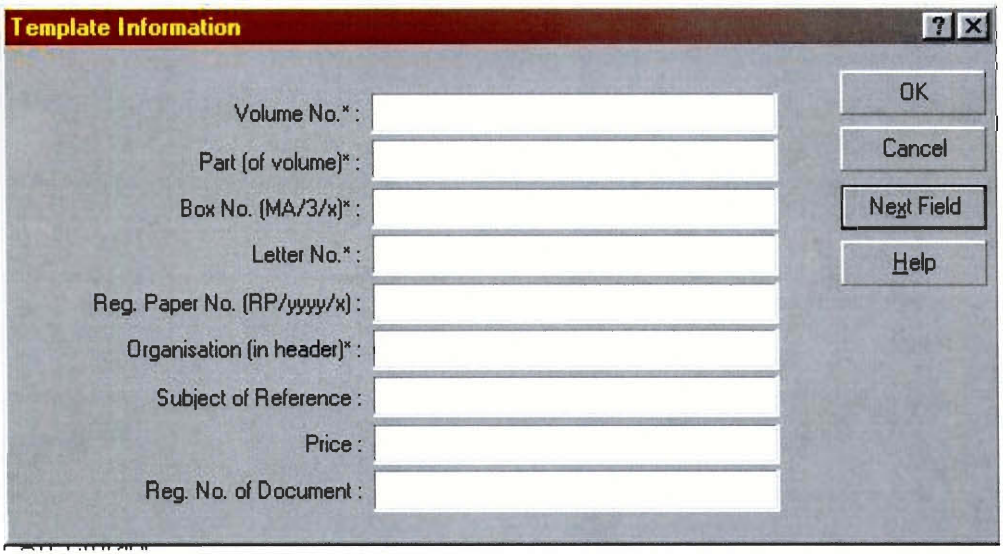

Figure 34 - Prompt dialog

Enter the information requested in its corresponding text box. It is acceptable to leave text box empty if the information asked for does not appear in the report.

***Note: Prompts that have an asterisk (*) next to their caption are required.

It is very important to note that sometimes not all of the prompts appear in the dialog box at once for a template. You need to scroll down through all of the prompts to see the ones that do not appear in the dialog box. This is a shortcoming of WordPerfect 9, as it does not indicate to the user that there are more prompts than can be seen. A good habit to avoid missing prompts is to press **enter** after completing each prompt; this will automatically move the cursor to the next prompt. If there is no information for a certain prompt, press enter without typing anything. This will make sure that you visit all of the prompts. When the last prompt is reached and the enter key is pressed, the dialog box will close and the information will be inserted into the template.

Note: The prompts associated with the individual templates will be displayed in ***bold italic to signify that they are unique to that template.

3.1.2 Common Template Structure

Each template has the same basic structure. The basic structure is made up of the following components:

- A teiHeader element, which contains bibliographic information about the source as well as the electronic transcription.
- A back section for information that sometimes appears on the back of the reports
- A main section that is uniquely structured for each template
- A section for adding minutes and annotations that appear on many of the reports
- A section for you to add comments and interpretations

The back section, the main section, the minutes and annotations section, and the comments section are all delineated by the or <div0> tag, which signifies that they are subdivisions of the document. The sections are distinguished using the tag when none of them contain any of their own subsections. When one of the sections contains its own subsections the <div0> tag is used to distinguish the three basic sections. Any subsections of the basic sections are then tagged using <div1> tags to signify another level of hierarchy. For example, if a document type contained columns, A and B, and you wanted to signify each column as a subsection of the document, the main section of the document would be structured as follows:

```
<div0>
      <div1 n=A type=column>…</div1>
      <div1 n=B type=column>…</div1>
</div0>
```

3.1.2.1 TeiHeader Element

Every document created will contain a teiHeader element at the top, which contains numerous subelements (See Figure 35). These tags are required in every document and contain important bibliographic information. You do not need to concern yourself with adding or editing any information in this element manually. You will be prompted for the information required to provide the bibliographic information for the report (For more information on prompts see section 3.1.1 *Prompts*).

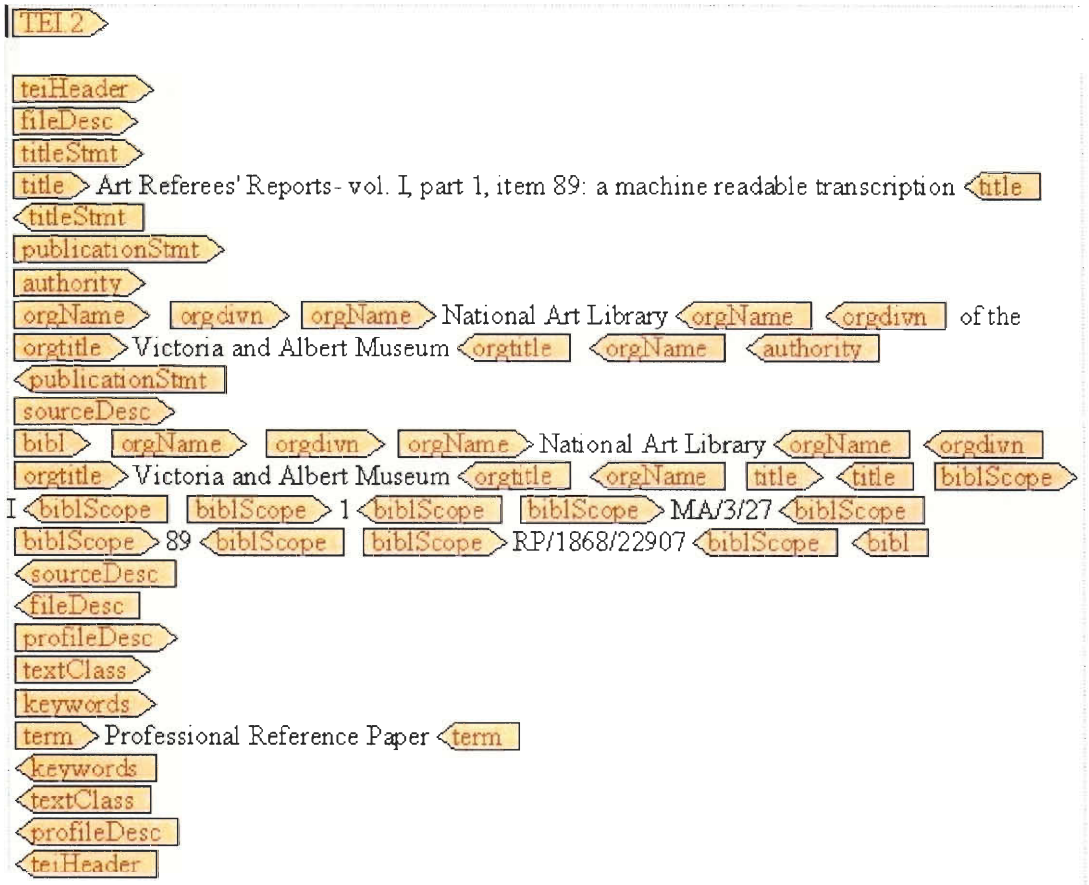

Figure 35 - teiHeader

3.1.2.2 Back Section

This section appears in all of the templates except the **Telegraph** template. This section is designated for the text that appears on the back of a report, excluding minutes and annotations. The back section is structured as follows:

```
Opener
      Date
Text
```

Opener

The <opener> tag is used to mark the heading of the back. When present, the opener is usually comprised of the date the letter was written.

Date

The <date> tag is used to mark the date the letter was written.

Text

The text can be a single paragraph or a series of paragraphs. Each paragraph must be marked using the paragraph tag, <p>. The text that usually appears on the back of a letter is a paragraph that briefly describes the content of the letter.

3.1.2.3 Main Section

After the teiHeader, each template has a main section that defines the unique structure and elements that are needed for that particular document type. This is where the majority of the actual transcribed text will be entered. The descriptions of individual templates will each have a section describing their unique main sections.

There are many reports that contain dates that have been provided as part of the report header that signify the date the report was written. It is important not to confuse this date will a date that appears on many of the pre-printed forms that is associated with

the stationary that the report is written on. The stationary date is usually pre-printed and appears in the upper-most top right hand corner of the report as shown in Figure 36.

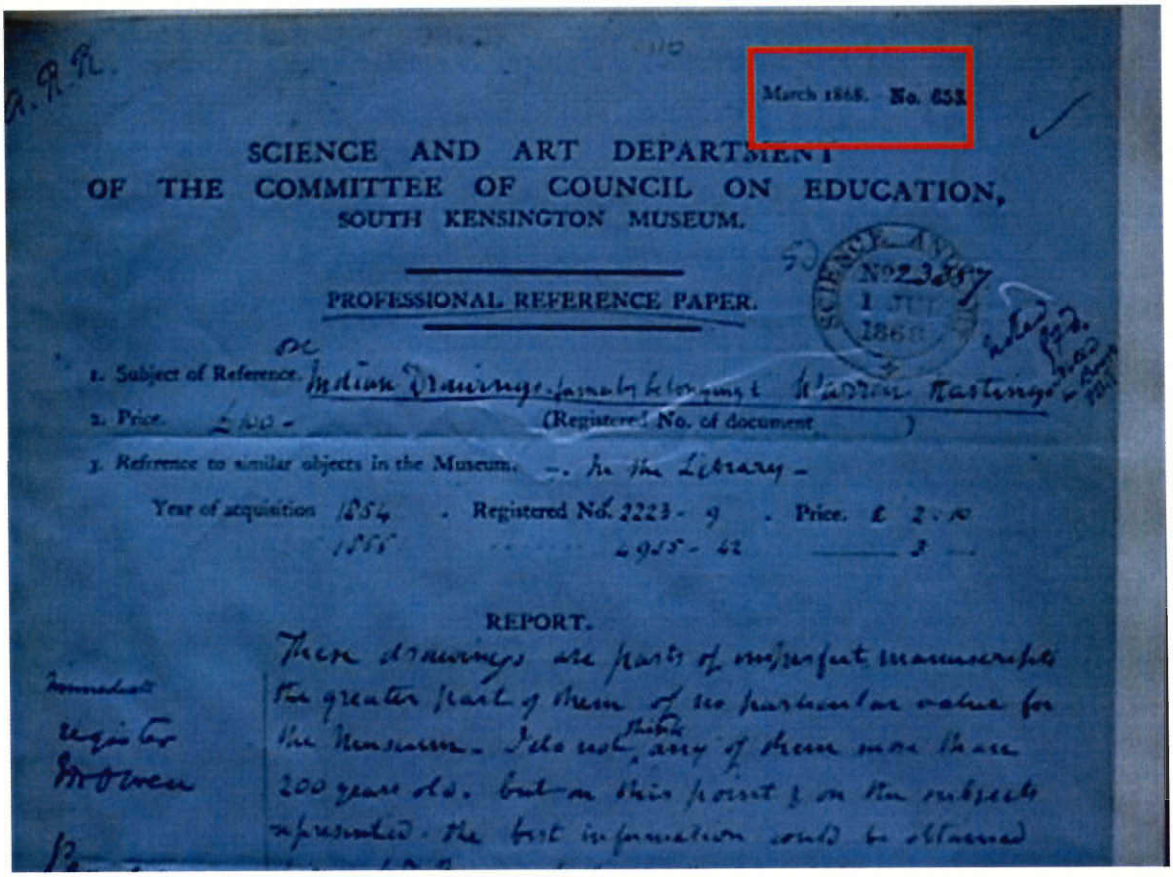

Figure 36 - Stationary Date

3.1.2.4 Minutes and Annotations

There are many minutes and annotations written by readers of the report that can be found on the reports themselves, as seen in Figure 37. They could be located pretty much anywhere on the document. Most of these minutes are generic to the whole document, or do not explicitly depict the part of the text that they refer to; a few,

however, refer to specific parts of the text, which are generally noted with a line or arrow.

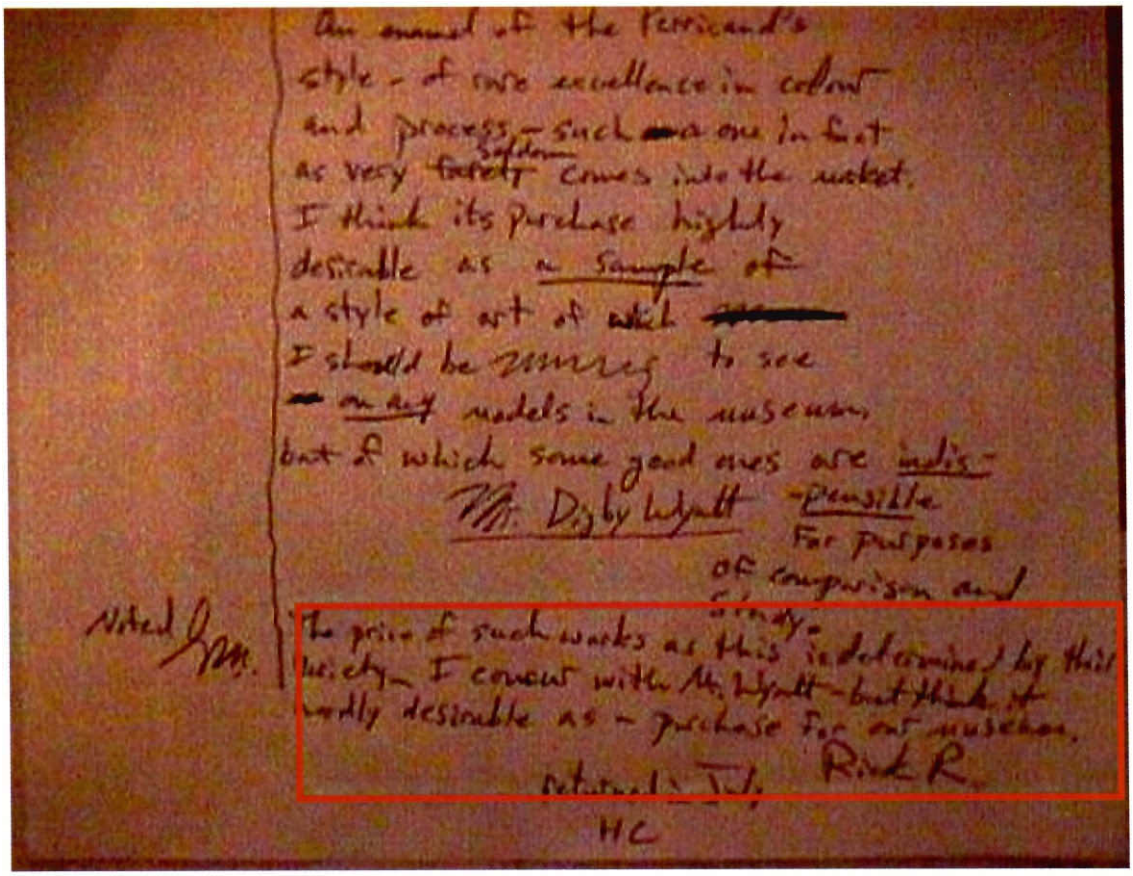

Figure 37 - Annotation

All minutes and annotations appearing on the report should be put in a <note> tag. The following information is associated with each note tag:

- The person responsible for writing it
- Its location on the document (head, foot, left, right, body, or footnote)

This information can be associated with the <note> tag by setting the resp and place attributes, respectively (For more information on setting attributes see section 2.5 *Editing an Attribute of a Tag*). The note shown in Figure 37 would be tagged as follows:

```
<note resp="Richard Redgrave" place=foot> The price of such works
as this is determined by their variety - I concur with Mr. Wyatt -
but think it hardly desirable as a purchase for our museum.
<persName>Rick R.</persName> </note>
```

***Note: If you do not know the full name of the person responsible, enter the information you know; otherwise enter *unknown*.

Each template, with the exception of minute paper, provides a section where all the general notes appear. The section appears as `Minutes and Annotations` in the editor as shown in Figure 38. Minute paper is treated differently because it is composed entirely of minutes and annotations. For information on how to handle notes on minute paper, see section 3.6 *The Minute* Paper.

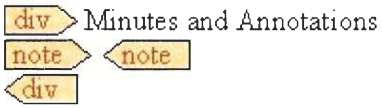

Figure 38 - Minutes and Annotations

It is **highly recommended that you copy the note tag already appearing in this section and paste a copy below the original for each minute or annotation beyond the first to preserve the default attribute settings**. If you insert a note tag from the element dialog it will not have the correct settings and you will need to set the attributes yourself.

For minutes and annotations that refer explicitly to specific areas of the text, footnotes can be inserted to make the association. Simply insert the footnote, as explained in section 2.6 *Inserting a Footnote*.

3.1.2.5 Comments

You may want to add comments to the transcription to offer insight or note some peculiarity. Your comments could include interpretations of the text, additional information about someone or something mentioned in the text, and remarks on the physical appearance of the report. Each template provides a section called `Comments` for

this purpose (See Figure 39). This section should also contain any references or relationships to other reports that you may find.

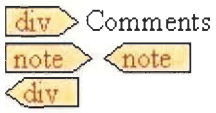

Figure 39 - Comments Section

Like minutes and annotations, each comment is contained within a <note> tag. However, there is no additional information that needs to be associated with the tag. The <note> tag in the comments section has all of its attribute values predefined.

Because of these predefined values, it is **highly recommended that you copy the <note> tag and paste it below the original for each additional comment that you wish to add**. If you insert a <note> tag from the element dialog it will not have the correct settings and you will need to set the attributes yourself.

However, if your comments are not generic to the entire document, you may wish to insert a footnote at the place of interest. As with the minutes and annotations, simply insert a footnote as explained in section 2.6 *Inserting a Footnote*.

3.2 The Addendum Template

The **Addendum** template is used to create a document of type printed materials and sketches and miscellaneous (For information on document types see Table 1 in section 1.4 *Document Templates*). An example of a report that would use this template is MA/3/10-280.

3.2.1 Prompts

The **Addendum** template has the following prompts:

- **Title** (required) – This is the title given by the National Art Library to the reports. The two titles are *J.C. Robinson's reports* for boxes MA/3/1 – MA/3/26 and *Art Referees' Reports* for boxes MA/3/27 – MA/3/51.

- **Volume No.** (required) – This prompt refers to the volume of which the letter is a part. This should be entered as a roman numeral. This information is located on the box, and can also be found in Appendix A.

- **Part (of volume)** (required) – Each volume is broken up into parts. This signifies what part of the volume, specified for the **Volume No.** prompt, the letter is located in. This information is located on the box, and can also be found in Appendix A.

- **Box No**. (required) – This prompt refers to the box number the letter is located in. This number must be entered in the form MA/3/x, where x is the box number. This number should appear on the box in the specified form; if it does not, it can be found in Appendix A. **Note: the box number does not refer to the white stickers on the boxes, such as 6A, 11B, etc.**

- **Letter No.** (required) – This prompt refers to the report number. This number is usually located at the top of each report, written in pencil. Great care must be taken when entering this number because there are often multiple numbers on the letters due to changes in the numbering system.

- **Reg. Paper No.** – This prompt refers to the registered paper number of the letter. This is located in a stamp that appears on almost all of the reports. This number must be entered in the form RP/yyyy/x, where yyyy is the year specified in the stamp and x is the number written in the stamp. For the stamp shown in Figure 40, the registered paper number would be entered as *RP/1863/6408*.

- **Back Date** – This prompt refers to the date that sometimes appears on the back of the report.

3.2.2 Main Section Structure

The main section of the **Addendum** template consists of a single paragraph or a series of paragraphs each denoted with the <p> tag. This template is mainly used for

reports that cannot be transcribed, such as sketches or printed materials. The main section of this template is used to describe the report in the words of the transcriber.

3.3 The Basic Letter Template

The **Basic Letter** template is used to create a document of type letter (For information on document types see Table 1 in section 1.4 *Document Templates*). Examples of letters that would use this template are MA/3/1-29 and MA/3/1-28.

3.3.1 Prompts

The **Basic Letter** template has the following prompts:

- **Title** (required) – This is the title given by the National Art Library to the reports. The two titles are *J.C. Robinson's reports* for boxes MA/3/1 – MA/3/26 and *Art Referees' Reports* for boxes MA/3/27 – MA/3/51.

- **Volume No.** (required) – This prompt refers to the volume of which the letter is a part. This should be entered as a roman numeral. This information is located on the box, and can also be found in Appendix A.

- **Part (of volume)** (required) – Each volume is broken up into parts. This signifies what part of the volume, specified for the **Volume No.** prompt, the letter is located in. This information is located on the box, and can also be found in Appendix A.

- **Box No**. (required) – This prompt refers to the box number the letter is located in. This number must be entered in the form MA/3/x, where x is the box number. This number should appear on the box in the specified form; if it does not, it can be found in Appendix A. **Note: the box number does not refer to the white stickers on the boxes, such as 6A, 11B, etc.**

- **Letter No.** (required) – This prompt refers to the report number. This number is usually located at the top of each report, written in pencil. Great care must be taken when entering this number because there are often multiple numbers on the letters due to changes in the numbering system.

- **Reg. Paper No.** – This prompt refers to the registered paper number of the letter. This is located in a stamp that appears on almost all of the reports. This number must be entered in the form RP/yyyy/x, where yyyy is the year specified in the

stamp and x is the number written in the stamp. For the stamp shown in Figure 40, the registered paper number would be entered as *RP/1863/6408*.

- **Back Date** – This prompt refers to the date that sometimes appears on the back of the report.
- ***Date*** – This prompt refers to the date that appears in the heading of the letter. **Note**: **Not all letters are dated.**

- ***Salutation*** – This prompt refers to the salutation of the letter, for example, *To Mr. Henry Cole*. **Note**: **Not all letters contain salutations.**

Figure 40 - Registered Paper Stamp

3.3.2 Main Section Structure

The **Basic Letter** template's main section is structured as follows:

```
Opener
      date
      salutation

Text of letter

Closer
      Signature
```

Opener

The <opener> tag is used to mark the opener of the letter. The opener contains all the heading information of the letter, such as the date and salutation.

Date

The <date> tag is used to mark the date the letter was written.

Salutation

The <salute> tag is used to mark the salutation of the letter and appears within the <opener> tag.

Text of letter

The text of the letter can be a single paragraph or a series of paragraphs. Each paragraph must be marked using the paragraph tag, <p>.

Closer

The <closer> tag is used to mark the closer of the letter. The closer contains the signature of the author and sometimes a closing salutation. A closing salutation tag is not provided by default because of its rare appearance. If a closing salutation is part of the letter, the salutation tag must be inserted manually.

Signature

The <signed> tag is used to mark the signature of the author. This tag appears within the <closer> tag.

3.4 The Form Report Template

The **Form Report** template is used to create a document of type Form Letter (For information on document types see Table 1 in section 1.4 *Document Templates*). An example of a report that would use this template is MA/3/46-97/1.

3.4.1 Prompts

The **Form Report** template has the following prompts:

- **Title** (required) – This is the title given by the National Art Library to the reports. The two titles are *J.C. Robinson's reports* for boxes MA/3/1 – MA/3/26 and *Art Referees' Reports* for boxes MA/3/27 – MA/3/51.

- **Volume No.** (required) – This prompt refers to the volume of which the letter is a part. This should be entered as a roman numeral. This information is located on the box, and can also be found in Appendix A.

- **Part (of volume)** (required) – Each volume is broken up into parts. This signifies what part of the volume, specified for the **Volume No.** prompt, the letter is located in. This information is located on the box, and can also be found in Appendix A.

- **Box No**. (required) – This prompt refers to the box number the letter is located in. This number must be entered in the form MA/3/x, where x is the box number. This number should appear on the box in the specified form; if it does not, it can be found in Appendix A. **Note: the box number does not refer to the white stickers on the boxes, such as 6A, 11B, etc.**

- **Letter No.** (required) – This prompt refers to the report number. This number is usually located at the top of each report, written in pencil. Great care must be taken when entering this number because there are often multiple numbers on the

letters due to changes in the numbering system.

- **Reg. Paper No.** – This prompt refers to the registered paper number of the letter. This is located in a stamp that appears on almost all of the reports. This number must be entered in the form RP/yyyy/x, where yyyy is the year specified in the stamp and x is the number written in the stamp. For the stamp shown in Figure 40, the registered paper number would be entered as *RP/1863/6408*.

- **Back Date** – This prompt refers to the date that sometimes appears on the back of the report.

- ***Date*** – This prompt refers to the date that appears in the heading of the letter.

- ***Salutation*** – This prompt refers to the opening salutation of the letter, for example, *Madame*. This text was often written in at the beginning of the printed form.

The next three prompts refer to the blanks left within the printed text for the author to fill in. These prompts are named by the one or more words appearing before the blank, followed by "...".

- ***visit ...*** - This prompt refers to the blank located in the first paragraph of the printed text.

- ***In charge of the ...*** - This prompt refers to the blank located in the second paragraph of the printed text.

- ***I am, ...*** - This prompt refers to the blank left in the closing salutation.

- ***To*** – This prompt refers to the `To` field located after the printed portion of the report but before the `Subjects for References` section.

3.4.2 Main Section Structure

The **Form Report** template's main section is made up of two subdivisions. The first subdivision is comprised of all the printed text. The second subdivision is the portion of the report marked `Subjects for References`. The main section is structured as follows:

```
Subdivision 1:

Opener
      date
      salutation

Text of letter
```

```
Closer
      salutation
      Signature
      To field


Subdivision 2:

      Text
```

Subdivision 1:

Opener

The <opener> tag is used to mark the opener of the letter. The opener contains all the heading information of the letter, such as the date and salutation.

Date

The <date> tag is used to mark the date the letter was written.

Salutation

The <salute> tag is used to mark the salutation of the letter and appears within the <opener> tag.

Text of letter

The text of the letter consists of three paragraphs. Each paragraph is marked using the paragraph tag, <p>.

Closer

The <closer> tag is used to mark the closer of the letter. The closer contains the signature of the author, the closing salutation, and the `To` field located at the end of the printed portion.

Salutation

The <salute> tag is used to mark the closing salutation of the letter. This tag appears within the <closer> tag.

Signature

The <signed> tag is used to mark the signature of the author. This tag appears within the <closer> tag.

Subdivision 2:

Text

The text appearing in this section can be a single paragraph or a series of paragraphs explaining the objects that need to be reviewed. Each paragraph must be marked using the paragraph tag, <p>.

You will notice that there are <hi> tags used throughout this template. These tags are used to signify the text that is printed on the report. Any text marked as printed will show up bold in the editor.

3.5 The List of Objects Template

The **List of Objects** template is used to create a document of type list (For information on document types see Table 1 in section 1.4 *Document Templates*). Some examples of reports that would use this template are MA/3/11-345 and MA/3/11-349.

3.5.1 Prompts

The **List of Objects** template has the following prompts:

- **Title** (required) – This is the title given by the National Art Library to the reports. The two titles are *J.C. Robinson's reports* for boxes MA/3/1 – MA/3/26 and *Art Referees' Reports* for boxes MA/3/27 – MA/3/51.

- **Volume No.** (required) – This prompt refers to the volume of which the letter is a part. This should be entered as a roman numeral. This information is located on the box, and can also be found in Appendix A.

- **Part (of volume)** (required) – Each volume is broken up into parts. This signifies what part of the volume, specified for the **Volume No.** prompt, the letter is located in. This information is located on the box, and can also be found in Appendix A.

- **Box No**. (required) – This prompt refers to the box number the letter is located in. This number must be entered in the form MA/3/x, where x is the box number. This number should appear on the box in the specified form; if it does not, it can be found in Appendix A. **Note: the box number does not refer to the white stickers on the boxes, such as 6A, 11B, etc.**

- **Letter No.** (required) – This prompt refers to the report number. This number is usually located at the top of each report, written in pencil. Great care must be taken when entering this number because there are often multiple numbers on the letters due to changes in the numbering system.

- **Reg. Paper No.** – This prompt refers to the registered paper number of the letter. This is located in a stamp that appears on almost all of the reports. This number must be entered in the form RP/yyyy/x, where yyyy is the year specified in the stamp and x is the number written in the stamp. For the stamp shown in Figure 40, the registered paper number would be entered as *RP/1863/6408*.

- **Back Date** – This prompt refers to the date that sometimes appears on the back of the report.

3.5.2 Main Section Structure

The **List of Objects** template's main section contains a header and a list. The header is comprised of any text appearing in the header of the report (See Figure 41). There are varying types of lists that can be found in the reports that fit this template. The list template should only be used for reports that are entirely comprised of a list. Reports that have lists located within the text should be transcribed with the appropriate template. The default list provided is a four-column *stable* list without labels (For information on lists see section 4.12 *Lists*). This default list structure can be modified to handle any list report by following the guidelines on setting up a list in section 4.12 *Lists*. The first two rows of the list are provided by default.

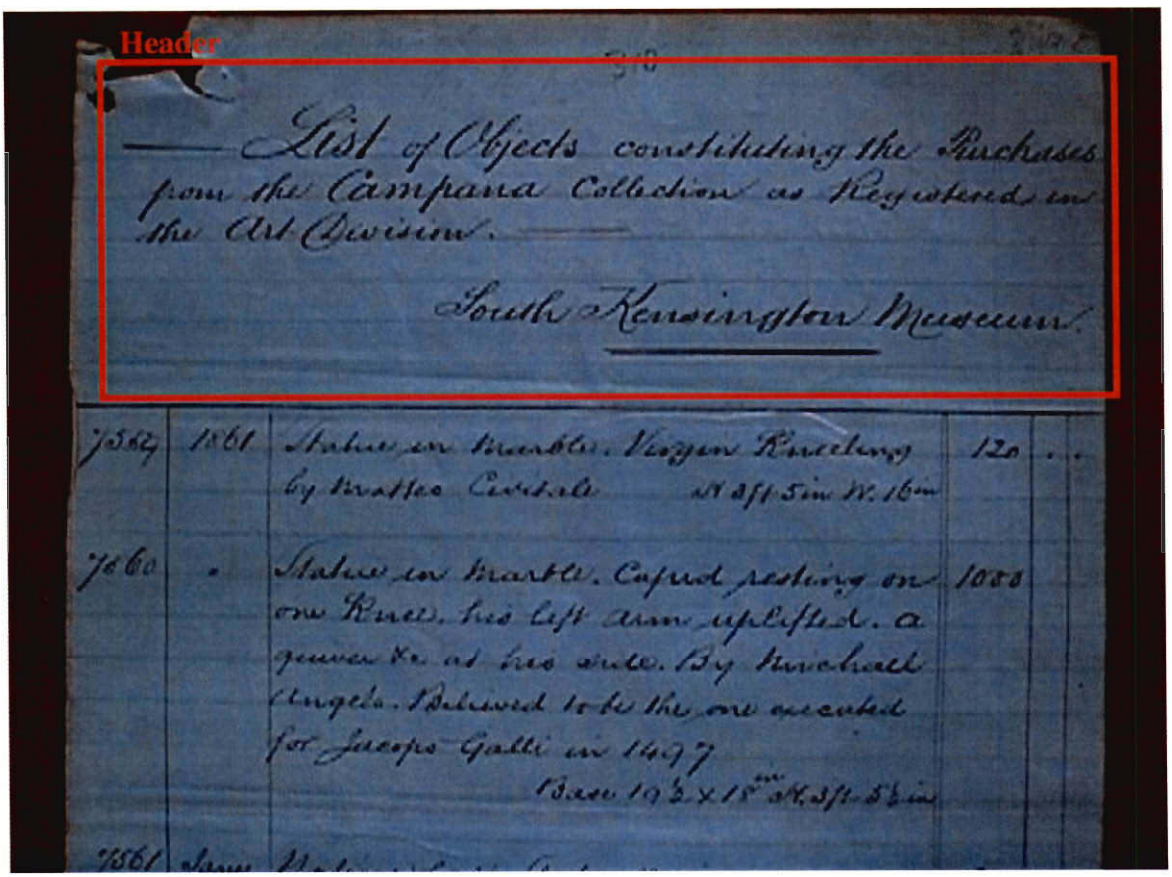

Figure 41 - List of Objects Report

To modify the list tag to make it a three-column list instead of the default four-column list the following steps would need to be taken:

1. Change the n attribute of the list to the value *3* (For information on setting attributes see section 2.5 *Editing an Attribute of a Tag*).

***Note: It is very important to complete this step because it will have a direct affect on how the list will be interpreted by the mechanism developed to display the report.

2. Remove an <item> tag from each "row." The default structure would need to be changed from:

```
<list n=4 type=stable>
      <item></item>
      <item></item>
      <item></item>
      <item></item>

      <item></item>
      <item></item>
      <item></item>
      <item></item>
</list>
```

To:

```
<list n=3 type=stable>
      <item></item>
      <item></item>
      <item></item>

      <item></item>
      <item></item>
      <item></item>
</list>
```

3.6 The Minute Paper Template

The **Minute Paper** template is used to create a document of type minute paper (For information on document types see Table 1 in section 1.4 *Document Templates*). Some examples of reports that would use this template are MA/3/12-16 and MA/3/46-71.

3.6.1 Prompts

The **Minute Paper** template has the following prompts:

- **Title** (required) – This is the title given by the National Art Library to the reports. The two titles are *J.C. Robinson's reports* for boxes MA/3/1 – MA/3/26 and *Art Referees' Reports* for boxes MA/3/27 – MA/3/51.

- **Volume No.** (required) – This prompt refers to the volume of which the letter is a part. This should be entered as a roman numeral. This information is located on the box, and can also be found in Appendix A.

- **Part (of volume)** (required) – Each volume is broken up into parts. This signifies what part of the volume, specified for the **Volume No.** prompt, the letter is located in. This information is located on the box, and can also be found in Appendix A.

- **Box No**. (required) – This prompt refers to the box number the letter is located in. This number must be entered in the form MA/3/x, where x is the box number. This number should appear on the box in the specified form; if it does not, it can be found in Appendix A. **Note: the box number does not refer to the white stickers on the boxes, such as 6A, 11B, etc.**

- **Letter No.** (required) – This prompt refers to the report number. This number is usually located at the top of each report, written in pencil. Great care must be taken when entering this number because there are often multiple numbers on the letters due to changes in the numbering system.

- **Reg. Paper No.** – This prompt refers to the registered paper number of the letter. This is located in a stamp that appears on almost all of the reports. This number must be entered in the form RP/yyyy/x, where yyyy is the year specified in the stamp and x is the number written in the stamp. For the stamp shown in Figure 40, the registered paper number would be entered as *RP/1863/6408*.

- **Back Date** – This prompt refers to the date that sometimes appears on the back of the report.

3.6.2 Main Section Structure

The **Minute Paper** template's main section contains a header and a series of minute note tags. The header contains one element, the museum name (shown in Figure 42), which is inserted in the <orgName> tag. Each note tag corresponds to a single

minute that appears on the report. These minutes are sometimes structured in an organized fashion, where the name, the date, and the minute will appear in their appropriate columns on the report. Other times, the minutes will appear all over the report in no organized way. Great care needs to be taken in choosing what text on the report constitutes as a single minute as well as the order in which they are entered. View the sample transcriptions of the minute paper to get a better grasp on how to identify minutes. It is **highly recommended that, if more than one minute appears on the report, the <note> tag provided by default with the template be copied and duplicated below the original.** This is recommended in order to preserve the default attribute settings of the <note> tag.

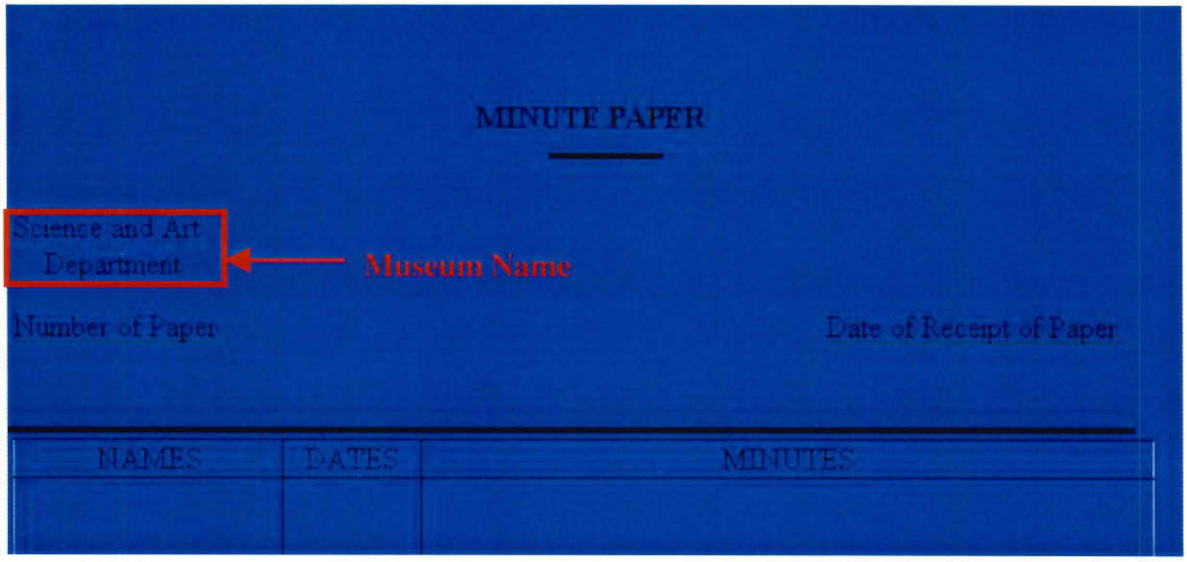

Figure 42 - Minute Paper Header

The **Minute Paper** template does not have a `Minutes and Annotations` section. This is because everything that appears in the minute paper report is considered to be a minute or annotation. Therefore, all minutes and annotations appear within the main section.

Like minutes on other types of reports, the minutes and annotations on minute paper must have the name of the person that wrote it and its location on the document specified in its resp and place attributes, respectively. For more information on the proper way of defining these attributes, see section 3.1.2.4 *Minutes and Annotations*.

In addition to the values available to the place attribute in normal minutes, minutes appearing on minute paper can have place defined as *body*. This value means that the minute is not an annotation, but part of the main body of the report.

3.7 The Telegraph Template

The **Telegraph** template is used to create a document of type telegraph (For information on document types see Table 1 in section 1.4 *Document Templates*). Some examples of reports that would use this template are MA/3/16-47, MA/3/16-50, and MA/3/23-216.

3.7.1 Prompts

The **Telegraph** template has the following prompts:

- **Title** (required) – This is the title given by the National Art Library to the reports. The two titles are *J.C. Robinson's reports* for boxes MA/3/1 – MA/3/26 and *Art Referees' Reports* for boxes MA/3/27 – MA/3/51.

- **Volume No.** (required) – This prompt refers to the volume of which the letter is a part. This should be entered as a roman numeral. This information is located on the box, and can also be found in Appendix A.

- **Part (of volume)** (required) – Each volume is broken up into parts. This signifies what part of the volume, specified for the **Volume No.** prompt, the letter is located in. This information is located on the box, and can also be found in Appendix A.

- **Box No**. (required) – This prompt refers to the box number the letter is located in. This number must be entered in the form MA/3/x, where x is the box number. This number should appear on the box in the specified form; if it does not, it can

be found in Appendix A. **Note: the box number does not refer to the white stickers on the boxes, such as 6A, 11B, etc.**

- **Letter No.** (required) – This prompt refers to the report number. This number is usually located at the top of each report, written in pencil. Great care must be taken when entering this number because there are often multiple numbers on the letters due to changes in the numbering system.

- **Reg. Paper No.** – This prompt refers to the registered paper number of the letter. This is located in a stamp that appears on almost all of the reports. This number must be entered in the form RP/yyyy/x, where yyyy is the year specified in the stamp and x is the number written in the stamp. For the stamp shown in Figure 40, the registered paper number would be entered as *RP/1863/6408*.

- ***Telegraph Company*** – This prompt refers to the telegraph company from which the telegraph was sent.

- ***Station*** – This prompt refers to the station name that sometimes appears on the telegraph.

- ***Date*** – This prompt refers to the date the telegraph was sent.

- ***Sender Name*** – This prompt refers to the name of the sender of the telegraph.

- ***Sender Address*** – This prompt refers to the address of the sender.

- ***Recipient Name*** – This prompt refers to the intended recipient of the telegraph.

- ***Recipient Address*** – This prompt refers to the recipient's address.

3.7.2 Main Section Structure

The **Telegraph** template's main section is structured as follows:

```
Header
      Telegraph Company Name
      Station Name
      Date
      Sender Name
      Sender Address
      Recipient Name
      Recipient Address

Telegraph message

Signature
```

Header

The <head> tag is used to mark the heading of the telegraph. The heading consists of the telegraph company name, the station name, the date, the sender and their address, and the recipient and their address.

Telegraph Company Name

The telegraph company name is marked with the <orgName> tag.

Station Name

The station name is marked with the <name> tag. The <name> tag has its type attribute set to *site* to signify that the name being marked is a name of a specific location.

Date

The <date> tag is used to mark the date the telegraph was written.

Sender Name

The <persName> tag with its type attribute set to *sender* is used to mark the name of the sender of the telegraph. This refers to the sender that appears in the *From* section of the telegraph.
***Note: The sender of the telegraph is not always entered as a name on the telegraph. For example, sometimes you will encounter in the *From* section of the telegraph a sender signified as *Secretary*.

Sender Address

The <addrLine> tag within the <address> tag is used to mark the address of the sender of the telegraph.

Recipient Name

The <persName> tag with its type attribute set to *recipient* is used to mark the name of the recipient of the telegraph. This refers to the recipient that appears in the *To* section of the telegraph.
***Note: The recipient of the telegraph is not always entered as a name on the telegraph. For example, sometimes you will encounter in the *To* section of the telegraph a sender signified as *Secretary*.

Recipient Address

The <addrLine> tag within the <address> tag is used to mark the address of the recipient of the telegraph.

Telegraph message

The text of the telegraph message can be a single paragraph or a series of paragraphs. Each paragraph must be marked using the paragraph tag, <p>.

Signature

The <signed> tag is used to mark the signature of the telegraph sender.

3.8 The ProfRefPaper Template

The **ProfRefPaper** template is used to create a document of type Professional Reference Paper (For information on document types see Table 1 in section 1.4 *Document Templates*). Some examples of reports that would use this template are MA/3/28-71 – MA/3/28-80.

3.8.1 Prompts

The **ProfRefPaper** template has the following prompts:

- **Volume No.** (required) – This prompt refers to the volume of which the letter is a part. This should be entered as a roman numeral. This information is located on the box, and can also be found in Appendix A.

- **Part (of volume)** (required) – Each volume is broken up into parts. This signifies what part of the volume, specified for the **Volume No.** prompt, the letter is located in. This information is located on the box, and can also be found in Appendix A.

- **Box No**. (required) – This prompt refers to the box number the letter is located in. This number must be entered in the form MA/3/x, where x is the box number. This number should appear on the box in the specified form; if it does not, it can be found in Appendix A. **Note: the box number does not refer to the white stickers on the boxes, such as 6A, 11B, etc.**

- **Letter No.** (required) – This prompt refers to the report number. This number is usually located at the top of each report, written in pencil. Great care must be taken when entering this number because there are often multiple numbers on the letters due to changes in the numbering system.

- **Reg. Paper No.** – This prompt refers to the registered paper number of the letter. This is located in a stamp that appears on almost all of the reports. This number must be entered in the form RP/yyyy/x, where yyyy is the year specified in the stamp and x is the number written in the stamp. For the stamp shown in Figure 40, the registered paper number would be entered as *RP/1863/6408*.

- **Back Date** – This prompt refers to the date that sometimes appears on the back of the report.

- ***Subject of Reference*** – This prompt refers to the `Subject of Reference` located in the header of the Professional Reference Paper.

- ***Price*** – This prompt refers to the `Price` located in the header of the Professional Reference Paper.

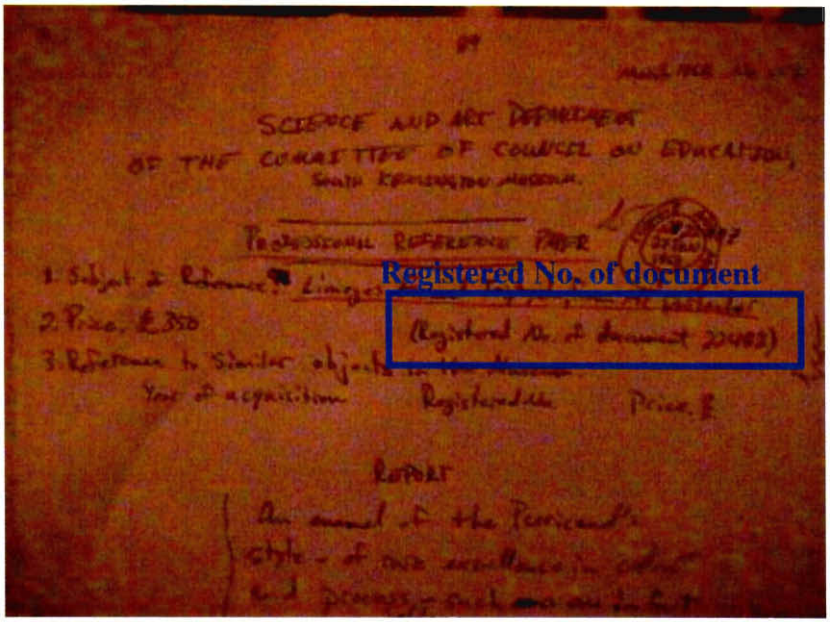

Figure 43 - Professional Reference Paper Information

- **Reg. No. of Document** – This prompt refers to the `Registered No. of document` located in the Professional Reference Paper (See Figure 43). This prompt should not be confused with the **Reg. Paper No.** prompt.

3.8.2 Main Section Structure

The **ProfRefPaper** template's main section is comprised of a header and two subdivisions, `Date and Country` and `Report`. The main section is structured as follows:

```
Header

Subsection 1: Date and Country
      Text

Subsection 2: Report
      Text
      Signature
```

Header

The <head> tag is used to mark the header of the Professional Reference Paper. The header is considered to be the portion signified in Figure 44. The header consists of the museum name, which you must fill in, and a single column ordered list (For information on lists see section 4.12 *Lists*). The first two items in the list are filled in by the prompts. The third item in the list is a three column *stable* list that contains information on similar objects in the museum.

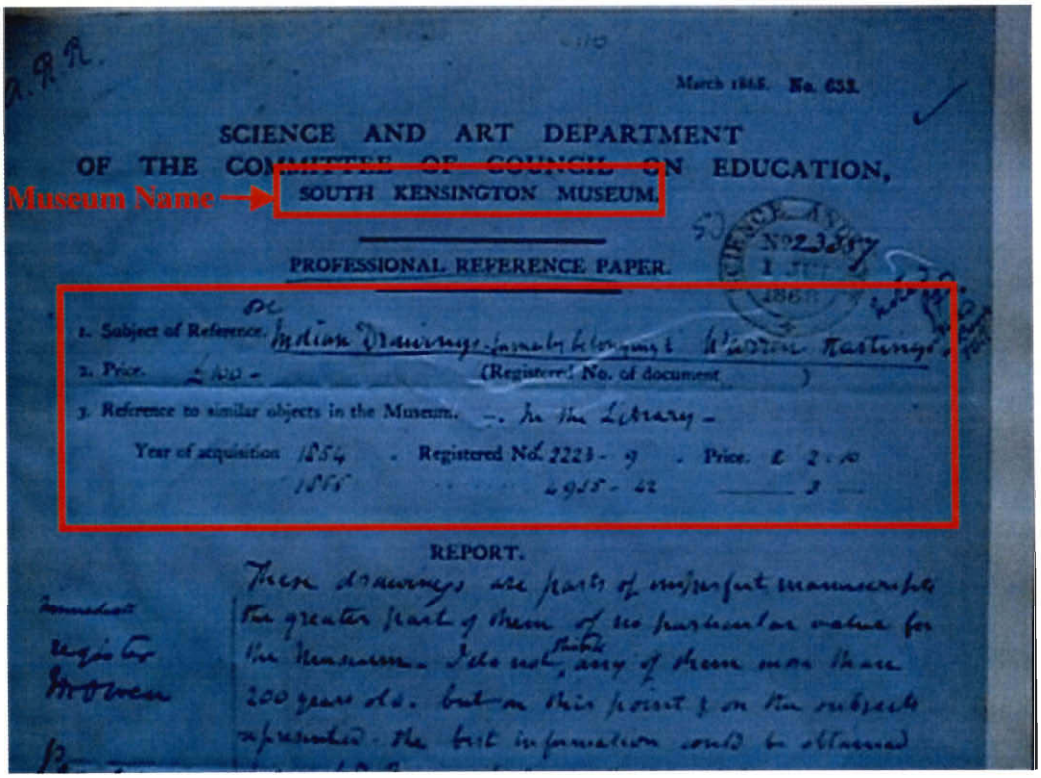

Figure 44 - Professional Reference Paper Header

The heading shown in Figure 44 would be marked up as shown in Figure 45:

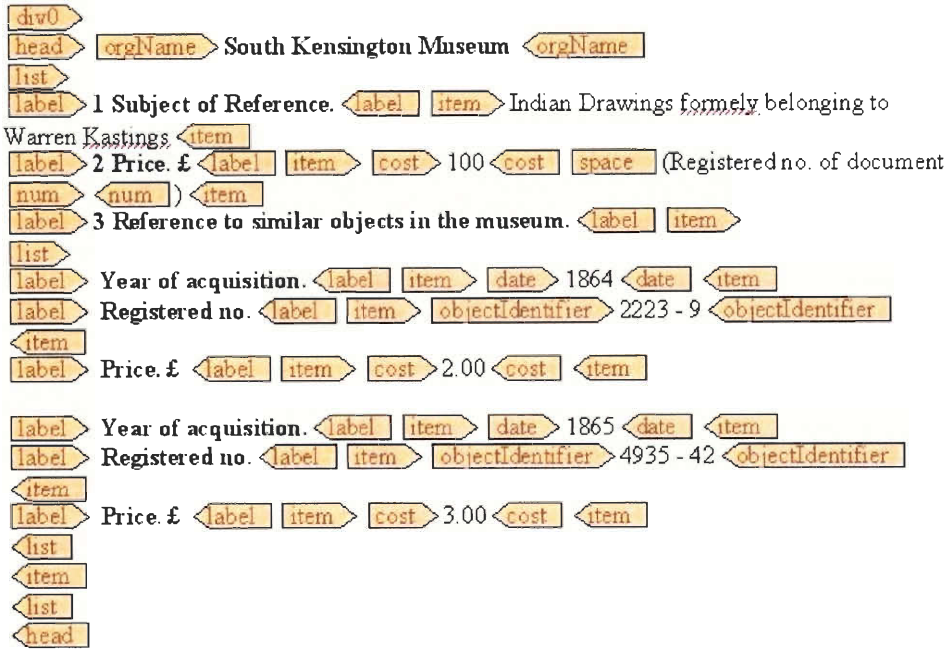

Figure 45 - Fully Marked up Professional Reference Paper Heading

It is **highly recommended that when more than one museum object is listed under item three that the label and item tags of the list be copied and pasted for each museum object listed.** The tags being

referred to are the label tags, containing the text `Year of acquisition, Registered no.,` and `Price £,` and their corresponding item tags. The copying is necessary to retain the special settings of the tags

Date and Country

This is the first subsection of the main portion of the report. This subsection corresponds to the first of the two columns that divides the report. It will often contain text signifying the date and origin of the piece being discussed in the Report section. Although this section is often labelled *Date and Country* on the report, it sometimes has no label.

***Note: this column is often used by readers of the report to make comments because it is generally left blank by the author. Special care should be taken when determining if text belongs in this subsection or it belongs in the minutes and annotations section.

Report

The Report section contains the message text and the signature of the author. The text can be a single paragraph or a series of paragraphs. Each paragraph must be marked using the paragraph tag, <p>. The signature is marked using the <signed> tag.

3.9 The ProfRefPaper(Rec From) Template

The **ProfRefPaper(Rec From)** template is a slight variation of the **ProfRefPaper** template. These reports contain an additional piece of information in the header: `Received from`. Some examples of reports that would use this template are MA/3/46-127 to MA/3/28-130.

This template consists of exactly the same prompts and structure as the **ProfRefPaper** template, with the exception of an additional prompt and an additional item in the ordered list in the header.

Prompt: ***Received From*** - This prompt refers to the `Received from` line located in the header of the Professional Reference Paper.

The additional item in the header's ordered list also corresponds to the `Received from` line in the header.

3.10 The Preliminary Reports Template

The **Preliminary Reports** template is used to create a document of type Preliminary Reports for the Information of the Director (For information on document types see Table 1 in section 1.4 *Document Templates*). An example of a report that would use this template is MA/3/27-47.

3.10.1 Prompts

The **Preliminary Reports** template has the following prompts:

- **Title** (required) – This is the title given by the National Art Library to the reports. The two titles are *J.C. Robinson's reports* for boxes MA/3/1 – MA/3/26 and *Art Referees' Reports* for boxes MA/3/27 – MA/3/51.

- **Volume No.** (required) – This prompt refers to the volume of which the letter is a part. This should be entered as a roman numeral. This information is located on the box, and can also be found in Appendix A.

- **Part (of volume)** (required) – Each volume is broken up into parts. This signifies what part of the volume, specified for the **Volume No.** prompt, the letter is located in. This information is located on the box, and can also be found in Appendix A.

- **Box No**. (required) – This prompt refers to the box number the letter is located in. This number must be entered in the form MA/3/x, where x is the box number. This number should appear on the box in the specified form; if it does not, it can be found in Appendix A. **Note: the box number does not refer to the white stickers on the boxes, such as 6A, 11B, etc.**

- **Letter No.** (required) – This prompt refers to the report number. This number is usually located at the top of each report, written in pencil. Great care must be taken when entering this number because there are often multiple numbers on the letters due to changes in the numbering system.

- **Reg. Paper No.** – This prompt refers to the registered paper number of the letter. This is located in a stamp that appears on almost all of the reports. This number must be entered in the form RP/yyyy/x, where yyyy is the year specified in the stamp and x is the number written in the stamp. For the stamp shown in Figure 40, the registered paper number would be entered as *RP/1863/6408*.

- **Back Date** – This prompt refers to the date that sometimes appears on the back of the report.

- ***Submitted to*** – This prompt refers to item 4 on the report, where it reads `Submitted to M` [Blank for entering a name] `for his remarks.` This prompt is for the person's name that appears in the blank.
  ***Note: The M, even though it is printed on the report needs to be included when filling in the prompt.

3.10.2 Main Section Structure

The **Preliminary Reports** template's main section is comprised of an ordered list, where each item in the list corresponds to an item printed on the report. For example, the first two items in the report are:

```
1. Description of object.
2. Price. £
```

Their corresponding items in the list are:

```
<label>1 Description of object.</label><item></item>
<label>2 Price. £</label><item><cost></cost></item>
```

The description of the object would be transcribed within the <item> tag and the price would be entered within its corresponding item tag. Since the price entered is a cost, it is tagged with the <cost> tag, which is provided by default in the template. The rest of the items follow a similar pattern, except for item 3. Item 3 is broken up into three parts on the report. Therefore, item 3 has an ordered list nested within its <item> tag, where a corresponding item in the list represents each of the three parts (See Figure 46).

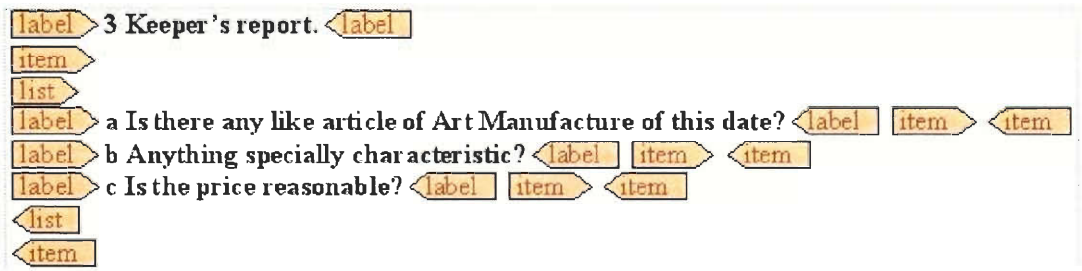

Figure 46 - Nested list

There is one important difference to note in the **Preliminary Reports** template about its overall structure; the report does not contain a `Back` section, because no reports of this type contain any text on the back.

3.11 Sample Transcriptions

Sample transcriptions have been provided for each of the document types in the Robinson Reports. These transcriptions were developed using the templates described earlier in this chapter. They are available in the folder **"C:\My Documents\Sample Transcriptions"**. For instructions on how to open the transcriptions, see section 2.1.3 *Opening a Document.*

The sample transcriptions are named with the same convention that you will be using. They follow the form *"MA-3-'Box No'-'Letter No'.xml"*. Each of the transcriptions can be related to a specific template, as seen in Table 2.

Table 2 - Sample transcription file names and templates

File Name	**Template**
MA-3-1-11.xml	Addendum
MA-3-1-3.xml	Basic Letter
MA-3-46-97-1.xml	Form Report
MA-3-11-349.xml	List of Objects (first page only)
MA-3-16-47.xml	Minute Paper
MA-3-46-71.xml	Minute Paper
MA-3-27-47.xml	Preliminary Reports
MA-3-27-49.xml	ProfRefPaper
MA-3-46-127.xml	ProfRefPaper(Rec From)
MA-3-16-53-3.xml	Telegraph

Replications of the original letters from which they were transcribed can be found in Appendix B. It is recommended that you use these transcriptions and sample letters as references when transcribing the Robinson Reports. They demonstrate the correct use of

the templates and possible uses of tags and attributes. Furthermore, you can practice transcribing the included letters to try and compare your results the sample transcriptions.

4 A Guide to the Tags

This chapter will discuss the different tags and attributes that you may used to mark up the transcriptions of the Robinson Reports. Before reading this chapter it is important that you understand several key concepts:

- Tags and tagging
- Empty tags
- Attributes
- Nesting

If you do not understand any of these concepts, refer to chapter 1 *Getting Started.*

Once you understand the underlying concepts behind tagging, you are ready to begin learning the tags themselves. Each tag will be given with the following information:

- A description of the tag
- Explanations of the attributes pertinent to the Robinson Reports
- Examples of the tags usage[3]

After you have read and fully understand the tags in this chapter, you will need to decide which ones *you* find pertinent. Most of these tags are not absolutely necessary; keep in mind, however, that they will be used:

- To provide formatting for the text
- To provide special display of certain types of text
- To facilitate searching for specific topics

If you feel that a tag is applicable to one of these areas, you should use the tag; otherwise, it is probably not as important. Attributes should be treated with the same discretion. An

[3] Many of these examples are taken directly from, or adapted from, the Robinson Reports; others are taken from the *TEI Guidelines for Electronic Text Encoding and Interchange.*

attribute should be used if it will help the formatting or display of the report, facilitate cataloguing and searching, or provide other useful information. If it does none of these things, the attribute can be ignored.

4.1 Global Attributes

Before you learn how to use the tags in the system, it is important that you understand some attributes that are common to all tags. These attributes, called *global attributes*, are present within every tag. They include:

- id Element Identifier
- **n Name or Number**
- **lang Foreign Words or Phrases**
- **rend Graphically Distinct Text**
- ana Analyses and Interpretations of Elements
- corresp Correspondence between Elements

***Note: You should only need to use those attributes marked in **bold** to complete the transcriptions.

***Note: In order to properly explain the global attributes, several tags are used in the examples. For explanations of these tags, see the appropriate sections of this chapter.

---

id Element Identifier

- id can be used to assign a unique identifier to any element – these ids *must* be unique throughout the entire collection
- The id must begin with a letter, but can use letters, numbers, hyphens, and periods

- Elements can be referred to by their id. For instance, it can be used to establish links and correspondence between different elements.

- Any attribute of type "IDREF" can refer to another element's id

Example:

*** **This I believe to be a poor galvano plastic reproduction of the reverse of a very fine medallion by Dupret of which the museum already possesses an admirable original.** ***

This example shows a paragraph, tagged , that has been given the id of p.1. This paragraph can now be referred to as p.1 by any other element.

***Note: Because of the complex nature of ids and id references, we recommend that you do not use the id attribute. This attribute can be assigned during a future pass through the transcriptions.

---

n Name or Number

- n is used to designate a name or number for a particular element

- n can be any string of characters

- You should only need to use n when creating lists. For information on creating lists, see section 4.12 *Lists*.

---

lang Foreign Words or Phrases

- lang designates text as being in a different language from the rest of the document

- If a segment of text is in a foreign language, and is marked by some tag, the lang attribute should be used to designate the language. If the segment is not marked by any tags, the <foreign> tag can be used with the lang attribute to designate the language (For information on <foreign>, see section 4.6 *Distinct Text*).

- When setting this attribute, use two letter abbreviations for languages, including:

    - fr for French
    - sp for Spanish
    - it for Italian
    - en for English
    - la for Latin
    - grc for Greek

    ***Note: If you need to add to this list, always remember to *be consistent* and *log your choices*

Example:

Meet me at the ***<emph lang=fr>****rendezvous****</emph>*** *point at midnight.*

This example shows how the tag that denotes emphasized text can specify that the text is french using the lang attribute. Remember, the lang attribute can be used within any tag. If you are noting a whole paragraph of text as foreign, you would use the lang attribute in the <p> tag.

---

rend Graphically Distinct Text

- rend is used to denote any text that is graphically distinct on the original document, such as underlined or printed text

- If a segment of text is graphically distinct, and is marked by some tag, the rend attribute should be used to designate the language. If the segment is not marked by any tags, the <hi> tag can be used with the rend attribute to designate the nature of the distinction (For information on <hi>, see section 4.6 *Distinct Text*).

- Acceptable values for rend are:

    - Underline
    - Quote (only used when the text is in quotation marks but is not an actual quotation)
    - Printed

***Note: You may add to this list, but always remember to *be consistent* and *log your choices*

Example:

Please call in ***<persName rend=underline>****Mr. Thompson****</persName>****.*

This example shows how the tag for a person's name can specify that the name is underlined using the rend attribute. Remember, the lang attribute can be used within any tag. If you are noting a whole paragraph of text as graphically distinct, you would use the rend attribute in the tag.

---

ana Analysis and Interpretations of Elements

- ana is used to link an element with any analyses or interpretations of that element
- ana must refer to a valid id attribute of another element

Example:

*** *** *This I believe to be a poor galvano plastic reproduction of the reverse of a very fine* ***<objectName id=o.1>****medallion****<objectName>*** *by Dupret of which the museum already possesses an admirable original.**** ***

...

<note ana=o.1>*This medallion could be the one that J.C. Robinson purchased for the museum in 1865****</note>***

This example shows how a note can be explicitly marked as an analysis of a particular element of the text by setting the ana attribute to the id of the element the note is interpreting.

***Note: Because of the complex nature of ids and id references, we recommend that you do not use the ana attribute. This attribute can be assigned during a future pass through the transcriptions.

---

corresp Correspondence between Elements

- corresp is used to show that two elements correspond in some way
- corresp must refer to a valid id attribute of another element

- corresp could be used to show relationships between different references to the same object by marking pronouns and antecedents. It can also be used to show relationships between different objects or passages that correspond in some way, such as an address and a sender or two identical passages written in different languages.

Examples:

*** *** *This I believe to be a poor galvano plastic reproduction of the reverse of a very fine* ***<objectName id=o.1>****medallion****<objectName>*** *by* ***<persName type=artist corresp=o.1>****Dupret****</persName>*** *of which the museum already possesses an admirable original.**** ***

This example shows how an artist's name can be marked as corresponding to a particular piece by setting the corresp attribute of the artist's <persName> tag to the id of the piece's <objectName> tag.

***Note: Because of the complex nature of ids and id references, we recommend that you do not use the corresp attribute. This attribute can be assigned during a future pass through the transcriptions.

4.2 Document Structure

The tags in this section determine document structure. Included in this section:

- ** Paragraph**
- **<pb> Page Break**
- **<lb> Line Break**

Paragraph

- Each paragraph in the Robinson Reports must be marked with
- Paragraphs do not need to be indented, as this formatting will be applied in the future

Example:

*** ****This I believe to be a poor galvano plastic reproduction of the reverse of a very fine medallion by Dupret of which the museum already possesses an admirable original.**** ***

<pb> Page Break

- <pb> can be inserted where there was a break in the original document
- <pb> can be inserted either within or between paragraph tags
- <pb> is an empty tag
- Page breaks are not necessary, unless you want to preserve the page breaks of the original text

Examples:

*** ****This I believe to be a poor galvano plastic reproduction of the reverse of a very fine medallion by Dupret of which the museum****<pb>****already possesses an admirable original.**** ***

*** ****I recommend for purchase of Mr. G. Eastwood a casket in brass or "latten" of the 15th century, engraved with animals, foliated ornamentation, and inscriptions in Gothic or church test characters. The museum does not possess any specimen of this class, which is a known and familiar one.**** ***
<pb>
*** ****And also of Mr. Whitehead a statuette of a child or "Amorino" in carved pear or boxwood on a pedestal of Boule work.**** ***

The first example shows that the previous paragraph spilled onto the next page between the words "museum" and "already."

The second example shows that there is a page break between the two paragraphs.

<lb> Line Break

- <lb> should be inserted where it is important to note that a line ended in the original document

- <lb> is an empty tag
- Line breaks are not necessary, unless you want to preserve the line breaks of the original text

Example:

*** ****I recommend for purchase of Mr. G. Eastwood****<lb>****a casket in brass or "latten" of the 15th****<lb>****century, engraved with animals, foliated****<lb>*** *ornamentation, and inscriptions in Gothic****<lb>****or church test characters. The museum****<lb>****does not possess any specimen of this class, which is a known and familiar one.**** ***

This example shows that the paragraph in the original text has five line breaks

4.3 Names

The tags in this section are used to tag different types of names. Included in this section:

- **<persName> Name of a Person**
- **<placeName> Name of a Place**
- **<orgName> Name of an Organisation**
- **<objectName> Name of an Object**
- **<collectionName> Name of a Collection**
- **<materialName> Name of a Material**
- **<eventName> Name of an Event**
- **<name type=site> Name of a Specific Site**

<persName> Name of a Person

- <persName> is used to tag a person's entire name, including surnames, forenames, titles, generational information, and so forth

- The <persName> tag has the attributes type and reg available to it

    - The type attribute specifies the type of person being tagged. Acceptable values for type are:

        - artist (an artist)
        - seller (an art seller)
        - previous owner (a previous owner of a piece)
        - reproducer (a reproducer of art, such as a photographer or caster)
        - author (the writer of the report)
        - recipient (the addressee of the report)
        - reader (a reader of the report, i.e. someone who signed it)

        - comrecipient (the addressee of a comment written onto the report by a reader)
        - other (any other person's name)

    ***Note: You may add to this list, but always remember to *be consistent* and *log your choices*

    - The reg attribute allows you to supply a regularised version of the name. This will probably be done completely, following set cataloguing rules, at a later date. However, if you can recognise a form of a name, such as initials, that others may not know, you may want to put the name in the reg attribute. The cataloguer could then re-enter the name in a standard form at a later date.

Examples:

<persName type=author>**J. C. Robinson**</persName>

<persName type=reader reg="Cole, Henry">**HC**</persName>

<persName>**John, by the grace of God, king of England, lord of Ireland, duke of Normandy and Aquitaine, and count of Anjou**</persName>

These examples show various forms of names that can be marked up. Notice the initials of Henry Cole are marked with <persName> with the full expansion of his name in the reg attribute. Also, note in the third example the extent to which a

name should be marked – all information that is considered part of a person's name in some way should be included.

<placeName> Name of a Place

- <placeName> indicates the name of a place, whether it be a geographic formation or a name of specific geo-political region

- <placeName> is used if the writer is referring to an area near a specific landmark, any distance or directional information relating the area of interest to the landmark

- <placeName> has the reg attribute available to it. The reg attribute allows you to supply a regularised version of the name. This will probably be done completely, following set cataloguing rules, at a later date. For the time being, though, you may want to tag an adjective with <placeName> and set the reg attribute to the standard name.

Examples:

<placeName>*three miles north of Lisbon****</placeName>***

<placeName>*Mount Everest****</placeName>***

<placeName reg=Great Britain>*GB****</placeName>***

<placeName reg=Finland>*Finnish****</placeName>***

<orgName> Name of an Organisation

- <orgName> is used to contain the name of an organisation of any sort, including private corporations, public organisations, and government agencies.

- The <orgName> tag has the attributes type and reg available to it

    - The type attribute specifies the type of organisation. Acceptable values for type are:

        - voluntary
        - political
        - governmental

- industrial
- commercial

***Note: You may add to this list, but always remember to *be consistent* and *log your choices*

- The reg attribute allows you to supply a regularised version of the name. This will probably be done completely, following set cataloguing rules, at a later date. You may, however, wish to use the reg attribute to tag phrases like *the museum* and signify which museum in particular is being referred to.

Examples:

<orgName>*The South Kensington Museum****</orgName>***

<orgName type=governmental reg="The Department of Science and Art">*The Department****</orgName>***

<orgName reg="Victoria and Albert Museum">*V&A****</orgName>***

<objectName> Name of an Object

- <objectName> denotes the name of an object, such as an artefact that Robinson was considering buying

- The <objectName> tag has the attributes type and reg available to it

  - The type attribute specifies the type of object. No values have been defined for this attribute, but if you choose to use it remember to *be consistent* and *log your choices*

  - The reg attribute allows you to supply a regularised version of the name. This will probably be done completely, following set cataloguing rules, at a later date. However, if you can recognise a form of a name that others may not know, you may want to put the name in the reg attribute. The cataloguer could then re-enter the name in a standard form at a later date.

Examples:

<objectName>*casket****</objectName>***

<objectName type=medallion reg="Medallion – Marriage of Henry IV"> *Medallion "Marriage of Henry IV"* ***</objectName>***

<collectionName> Name of a Collection

- <collectionName> is used to mark the name of a collection, as opposed to individual pieces
- The <collectionName> tag has the attributes type and reg available to it
    - The type attribute specifies the type of collection. No values have been defined for this attribute, but if you choose to use it remember to *be consistent* and *log your choices*
    - The reg attribute allows you to supply a regularised version of the name. This will probably be done completely, following set cataloguing rules, at a later date. However, if you can recognise a form of a name that others may not know, you may want to put the name in the reg attribute. The cataloguer could then re-enter the name in a standard form at a later date.

Examples:

<collectionName>*a collection of lace****</collectionName>***

<collectionName type=personal reg="Collection – Wilson">*the Wilson Collection****</collectionName>***

<materialName> Name of a Material

- <materialName> signifies the material of which an object is made
- The <materialName> tag has the attributes type and reg available to it
    - The type attribute specifies the type of material. No values have been defined for this attribute, but if you choose to use it remember to *be consistent* and *log your choices*
    - The reg attribute allows you to supply a regularised version of the name. This will probably be done completely, following set cataloguing rules, at

a later date. However, if you can recognise a form of a name that others may not know, you may want to put the name in the reg attribute. The cataloguer could then re-enter the name in a standard form at a later date.

Examples:

<materialName type=metal>*brass****</materialName>***

<materialName type=wood reg=pine> *evergreen****</materialName>***

<eventName> Name of an Event

- <eventName> specifies the name of a specific event
- The <eventName> tag has the attributes type and reg available to it
    - The type attribute specifies the type of event. No values have been defined for this attribute, but if you choose to use it remember to *be consistent* and *log your choices*
    - The reg attribute allows you to supply a regularised version of the name. This will probably be done completely, following set cataloguing rules, at a later date. However, if you can recognise a form of a name that others may not know, you may want to put the name in the reg attribute. The cataloguer could then re-enter the name in a standard form at a later date.

Examples:

<eventName type=exhibition>*Loan Exhibition****</eventName>***

<eventName type=holiday reg="Independence Day">*the fourth of July****</eventName>***

<name type=site> Name of Specific Site

- <name>, with the type attribute set to *site* is used to indicate specific locations that are not considered organisations, such as telegraph stations

- There is no default type attribute for <name> – you will need to set the type attribute to *site* every time you use this tag

- This tag should *only* be used to signify a site, as other types of names are covered by the more specific name tags

- The reg attribute is available to <name type=site>. The type attribute allows you to supply a regularised version of the name. This will probably be done completely, following set cataloguing rules, at a later date. However, if you can recognise a form of a name that others may not know, you may want to put the name in the reg attribute. The cataloguer could then re-enter the name in a standard form at a later date.

Examples:

<name type=site>*Charing Cross Station****</name>***

<name type=site reg="Charing Cross Station">*Ch. Cr. Sta.****</name>***

4.4 Dates

The tags in this section are used to tag dates. Included in this section:

- **<date> A Date**
- **<dateRange> A Range of Dates**

<date> A Date

- <date> is used to mark any manner of date
- The <date> tag has the attributes value and certainty available to it
    - The value attribute specifies the date in the standard form yyyy-mm-dd. If the day or month is not known, the form is shortened to either yyyy-mm or simply yyyy.
    - The certainty attribute is used in the <date> tag to indicate how certain you are that the date is precise. The value of certainty should be a number between 0 and 100, with 0 meaning that the date is extremely imprecise and 100 meaning that the date is completely precise.

Examples:

<date value=1865-04-14 certainty=100>*14 April 1865****</date>***

<date value=1871>*1871****</date>***

<date value=1868-06-12 certainty=75>*the twelfth of June, 1868****</date>***

<dateRange> A Range of Dates

- <dateRange> is used to tag a range of dates
- The <dateRange> tag has the attributes from, to, and exact available to it
    - The from and to attributes specify the starting and ending point of the period in the standard form yyyy-mm-dd. If the day or month are not known, they are dropped and the form becomes either yyyy-mm or simply yyyy.
    - The exact indicates how precise the dates given in from and to are. Legal values for exact are:
        - from (the from date is exact)
        - to (the to date is exact)
        - both (both dates are exact)
        - none (neither date is exact)

Examples:

<dateRange from=1863 to=1886 exact=both>*1863-1886****</dateRange>***

<dateRange from=1450 to=1500>*the second half of the 15th Century****</dateRange>***

<dateRange from=1740 to=1760 exact=none>*circa 1750****</dateRange>***

4.5 Numbers

The tags in this section are used to tag numbers. Included in this section:

- **<objectIdentifier> Museum Object Number**
- **<cost> Cost of a Purchase**
- **<measure> A Quantity**
- **<num> A Number**

<objectIdentifier> Museum Object Number

- <objectIdentifier> is used to tag a Museum Object Number
- In Robinson's time, Museum Object Numbers took the form: 345-64, where 64 was short for the year
- The reg attribute can be used in <objectIdentifier> to regularise the Museum Object Number. These numbers now use the four-digit year instead of the two-digit year.

Examples:

<objectIdentifier>*345-64****</objectIdentifier>***

<objectIdentifier reg=182-1871>*182-71****</objectIdentifier>***

<cost> Cost of a Purchase

- <cost> denotes the price of an object that Robinson was reporting on
- This could be the requested price or the price actually paid

Examples:

*The price is £****<cost>****5.0.0****</cost>***

<measure> A Quantity

- <measure> refers to any quantity of a thing
- <measure> usually contains a number, a unit, and the name of the thing being measured
- The <measure> tag has the attributes type, value, unit, and reg available to it
    - The type attribute in measure specifies the type of unit the measurement is taken in. It can be set to one of the following:
        - Weight
        - Count (dozen, score)
        - Length
        - Area
        - Volume

    ***Note: You may add to this list, but always remember to *be consistent* and *log your choices*

    - The value attribute is a number that specifies the amount of the thing being measured.
    - The unit attribute specifies the unit that is being used in the measurement.
    - The reg attribute specifies a regularised version of the measurement, usually recorded in the metric system. This is generally given with the value and unit; for example, *70cm*.

Examples:

<measure type=length>*seventy centimetres****</measure>***

<measure type=volume value=12 unit=gallon reg=54.5l>*twelve gallons of milk****</measure>***

<num> A Number

- <num> refers to any form of a number
- It is not necessary to mark *every* number; only those that you feel are somehow significant should be marked.

- Attributes for <num> include type and value

    - The type attribute specifies the type of number being marked. It can be set to one of the following:

        - Cardinal (5, V, five)
        - Ordinal (5th, fifth)
        - Fraction (1/5, .2, one fifth)
        - Percentage (5%, five percent)

    ***Note: You may add to this list, but always remember to *be consistent* and *log your choices*

    - The value attribute is used to regularise the value of a number

Examples:

<num type=cardinal value=1868>**MDCCCLXVIII**</num>

<measure type=length><num type=cardinal value=70>**seventy**</num> *centimetres****</measure>***

These two examples show how the <num> tag can be used to mark any type of number.

The second example shows how <num> can be nested within other tags. Here, it is nested within a measurement of seventy centimetres, marked as a measure of length, with the number "seventy" marked as a cardinal number with the value 70.

4.6 Distinct Text

The tags in this section are used to tag text that is linguistically or graphically different than the surrounding text. Included in this section:

- **<quote> A Quotation**
- **<hi> Graphically Distinct Text**
- **<emph> Emphasized Text**
- **<distinct> Distinct Words or Phrases**

- **<foreign> Foreign Words or Phrases**
- **<soCalled> Unusual Words or Phrases**
- **<mentioned> Mentioned Words or Phrases**
- **<space> Blank Space**

<quote> A Quotation

- <quote> marks a quotation attributed to an outside source
- The editor you are using cannot interpret quotation marks, so *always* marks quotations with the <quote> tag
- Paragraph tags cannot be inserted within <quote>. If a quotation spans several paragraphs, each paragraph must be marked separately as a <quote>

Examples:

As Robinson said in one of his reports, ***<quote>****The museum does not possess any specimen of this class, which is a known and familiar one.****</quote>***

Robinson went on to say:
*** <quote>****I recommend for purchase of Mr. G. Eastwood a casket in brass or "latten" of the 15th century, engraved with animals, foliated ornamentation, and inscriptions in Gothic or church test characters. The museum does not possess any specimen of this class, which is a known and familiar one.****</quote> ***
*** <quote>****And also of Mr. Whitehead a statuette of a child or "Amorino" in carved pear or boxwood on a pedestal of Boule work.****<quote> ***

***Note: The quotation marks used on Amorino and latten must be replaced with either the <hi> tag or the <soCalled> tag described later in this section. Do not use the <quote> tag in areas like these, as they are not quotations.

<hi> Graphically Distinct Text

- <hi> is used to denote any text that is graphically distinct on the original document, such as underlined text.

- The rend attribute in <hi> is used to describe the nature of the distinction. It can be set to one of the following:

    - Underline
    - Quote (Note: <quote> should be used for quotations. Other quotation marks *must* be replace with either <hi> or <soCalled>, discussed below)
    - Printed

***Note: You may add to this list, but always remember to *be consistent* and *log your choices*

Examples:

a child or "Amorino"
would be tagged
a child or ***<hi rend=quote>****Amorino****</hi>***

desirable as a sample of
would be tagged
desirable as ***<hi rend=underline>****a sample****</hi>*** *of*

<emph> Emphasized Text

- <emph> signifies text that is stressed or emphasized linguistically

Example:

the interiors of the architectural monuments of Spain are ***<emph>****everything* ***</emph>*** *and the exteriors of little moment*

<distinct> Distinct Words or Phrases

- <distinct> is used to mark linguistically distinct words or phrases, such as slang, technical or archaic terms.

- Attributes for <distinct> include type, time, space, and social

    - The type attribute specifies the sublanguage to which the word or phrase belongs, such as archaic, technical, or slang.

    - The time attribute specifies how the text is distinct due to time, such as archaic, old-fashioned, contemporary, or futuristic.

    - The space attribute specifies how the text is distinct due to location, such as English, African, or Texan.

    - The social attribute specifies how the text is distinct due to social classifications, such as technical, polite, impolite, or restricted.

    ***Note: No guidelines exist for what these attributes should be set to. If you choose to use them, remember to *be consistent* and *log your choices.* Keep in mind, however, that none of these attributes are necessary.

Example:

Next morning a boy in that dormitory confided to his bosom friend, a ***<distinct time=1900 social=publicschool space=gb>****fag****</distinct>*** *of Macrea's, that there was trouble in their midst which King* ***<distinct type=archaic>****would fain****</distinct>*** *keep secret.*

This example shows how the four available attributes are used. The word "fag" is marked as being from public schools in Great Britain in the year 1900. The phrase "would fain" is marked more simply as being archaic.

<foreign> Foreign Words or Phrases

- <foreign> designates text as being in a different language from the rest of the document

- <foreign> is used when the word or phrase is not otherwise tagged. If it is already tagged as something else, use the lang attribute in that tag to designate the language (see section 4.1 *Global Attributes*)

- Use the lang attribute in the <foreign> tag to specify the language. When setting this attribute, use two letter abbreviations for languages, including:

    - fr for French
    - sp for Spanish
    - it for Italian

- en for English
- la for Latin
- grc for Greek

***Note: If you need to add to this list, always remember to *be consistent* and *log your choices*

Example:

Meet me at the ***<foreign lang=fr>****rendezvous****</foreign>*** *point at midnight.*

This example shows the word *rendezvous* being marked as French. However, if the word was meant to appear in quotes in the final display, it would be more appropriate to mark it with <hi> and set its lang attribute to French, as shown here:

Meet me at the ***<hi rend=quote lang=fr>****rendezvous****</hi>*** *point at midnight.*

Likewise, if you were noting a whole paragraph of text as foreign, you would use the lang attribute in the <p> tag. Only use the <foreign> tag if the foreign text is not otherwise marked.

<soCalled> Unusual Words or Phrases

- <soCalled> is used to designate words or phrases that are not normally used and the author does not take credit for
- <soCalled> may be used for words in "scare quotes"

Example:

a child or ***<soCalled>****Amorino****</soCalled>***

***Note: Because the editor cannot interpret quotation marks, either <soCalled> or <hi> *must* be used for scare quotes

---

<mentioned> Mentioned Words or Phrases

- <mentioned> marks a word or phrase that is mentioned in the text, but is not used

Example:

There is thus a striking accentual difference between a verbal form like ***<mentioned>****eluthemen****</mentioned>*** *accented on the second syllable of the word, and its participial derivative* ***<mentioned>****lutheis****</mentioned>*** *accented on the last.*

<handShift> Change in Hand

- <handShift> is used to mark a change in hand, writing style, or ink
- <handShift> is an empty tag
- The <handShift> tag has the attributes character, ink, and style available to it
    - The character attribute is used to describe the quality of the writing. These could include *sloppy* and *neat.*
    - The style attribute is used to denote recognised writing styles, such as *anglicana* or *secretary.*

    ***Note: Neither character nor style has a pre-defined set of acceptable values. If you choose to use these attributes, you will need to choose values for them. Remember, always *be consistent* and *log your choices.*
    - The ink attribute describes the colour of the ink being used. This should be either a colour word, like *red*, *blue,* or *brown,* or the word *pencil* to describe text written in pencil.

Example:

<handShift ink=black>*I think its purchase highly desirable as a sample of the style of art of****<handShift ink=grey>****which I should be willing to see on any models in the museum.*

This example shows a paragraph that started off in black ink, but switched to grey half way through. No change in the characteristics or style of the writing was noted.

<space> Blank Space

- <space> is used to designate a noticeably large amount of space in a line of text
- <space> is an empty tag
- The <space> tag has the attributes dim and extent available to it
    - The dim attribute specifies the direction of the space. Legal values for dim are:
        - vertical
        - horizontal
    - The extent attribute gives the approximate amount of space. This can be given in any units, but the units must be specified. For example, *1cm* or *three characters*. As always, remember to *be consistent* and *log your choices*.

Examples:

*** ****I recommend for purchase of Mr. G. Eastwood a casket in brass or "latten" of the 15th century, engraved with animals, foliated ornamentation, and inscriptions in Gothic or church test characters. The museum does not possess any specimen of this class, which is a known and familiar one.**** ***
<space dim=vertical extent=5cm>
*** ****And also of Mr. Whitehead a statuette of a child or "Amorino" in carved pear or boxwood on a pedestal of Boule work.**** ***

*** ****This I believe to be a poor galvano plastic reproduction of the reverse of a very fine medallion by Dupret of which the museum****<space type=horizontal extent="three characters">*** *already possesses an admirable original.**** ***

The first example shows that there is five centimetres of space between the two paragraphs. The second shows a horizontal space of three characters between the words museum and already.

4.7 Illegible or Damaged Text

The tags in this section are used to tag text that is partially or completely illegible.

Included in this section:

- **<unclear> Unclear Text**

- **<damage> Damaged Text**
- **<gap> Unreadable Text**
- **<supplied> Text Supplied by the Transcriber**

<unclear> Unclear Text

- <unclear> is used to indicate text that is not fully legible and cannot be transcribed with absolute confidence.
- The cert attribute is used in the <unclear> tag to indicate how certain you are that the transcribed text is correct. The value of certainty should be a number between 0 and 100, with 0 meaning that the text is definitely wrong and 100 meaning that the text is definitely correct.

  ***Note: If the certainty is 0, the <supplied> tag described later in this section should be used in place of <unclear>. If the certainty is 100, no tag is needed.

Example:

*** **And also of Mr. Whitehead a statuette of a child or **<unclear cert=85>**"Amorino"**</unclear>** in carved pear or boxwood on a pedestal of **<unclear cert=70>**Boule**</unclear>** work.** ***

This example shows that the transcriber was only 85% certain that "Amorino" was the correct word, and was only 70% certain that "Boule" was correct.

***Note: If the text cannot be read at all, the <gap> tag, described later in this section, should be used.

<damage> Damaged Text

- <damage> signifies text that cannot be transcribed with certainty due to damage of any kind

- The <damage> tag has the attributes type, extent, and degree available to it

    - The type attribute should be used to describe the type of damage. Valid values for type are:

        - inkblot
        - faded
        - burnt
        - torn
        - stained

      Note: You may add to this list, but always remember to *be consistent* and *log your choices*

    - The degree attribute defines the degree to which the text is damaged. Values for degree should range between 0 and 100, with 0 representing no damage and 100 representing damaged to full illegibility.

      ***Note: If the degree is 0, no tag is necessary; if it is 100, the <damage> tag must be used in conjunction with <gap> tag described below.

    - The extent attribute gives the amount that the text is damaged. This can be given in any units, but the units must be specified. For example, *one word* or *3cm*. As always, remember to *be consistent* and *log your choices*.

Example:

*** **The price of such works as this **<damage type= "inkblot" degree=40 extent="two words">**is determined**</damage>** by their variety – I concur with Mr. Wyatt – but think it hardly desirable as a purchase for our museum.** ***

This example shows that the words "is determined" are damaged by an inkblot. The degree of the damage is 40% and the extent of the damage is approximately 2 words.

<gap> Unreadable Text

- <gap> indicates where you cannot make out a given portion of text well enough to make a guess at what it says

- <gap> is an empty tag

- The <gap> tag has the attributes reason, extent, and desc available to it

    - The reason attribute gives the reason that the text has been omitted. Values for reason include:

        - illegible
        - damaged
        - cancelled

    ***Note: You may add to this list, but always remember to *be consistent* and *log your choices*

    - The extent attribute gives the amount that the text is damaged. This can be given in any units, but the units must be specified. For example, *one word* or *three characters*. As always, remember to *be consistent* and *log your choices*.

    - The desc attribute can be used to describe the type of information that is unreadable. For instance, if it is an illegible signature, desc could be set to *persName*. There are no defined values for this attribute, so if you choose to use it, remember to *be consistent* and *log your choices*.

Examples:

I think its purchase highly desirable as a sample of the style of art of which ***<gap reason=cancelled extent="one word">*** *I should be* ***<gap reason=illegible extent="three characters">*** *to see* ***<gap reason=damaged extent="5cm">*** *on any models in the museum.*

This example shows three gaps in the transcription. One item is crossed-out, one is illegible, and one is damaged.

If you wish to describe the type of damage that befell the third word, <gap> can be nested within a <damage> tag, as seen below:

I think its purchase highly desirable as a sample of the style of art of which ***<gap>*** *I should be* ***<gap>*** *to see* ***<damage type=inkblot><gap extent=1></damage>*** *on any models in the museum.*

Notice that the reason attribute in <gap> is not necessary if it is nested within a <damage> tag.

Note: If the illegible, damaged, or cancelled text can be read, or partially read, the <unclear>, <damage>, and <del> tags should be used respectively.

<supplied> Text Supplied by the Transcriber

- <supplied> can be used in place of <gap> if you wish to enter invented text in place of original text that is completely illegible

- The attribute reason is used in <supplied> to describe why text needed to be supplied. Values for reason include:

    - illegible
    - damaged
    - cancelled

***Note: You may add to this list, but always remember to *be consistent* and *log your choices*

Example:

I think its purchase highly desirable as a sample of the style of art of which ***<supplied reason=cancelled>****I would like****</supplied>*** *I should be* ***<supplied reason=illegible>****willing****</supplied>*** *to see* ***<supplied reason=damaged>****used****</supplied>*** *on any models in the museum.*

This example shows how text can be supplied to replace that which cannot be read.

As with the <gap> tag, you can describe the type of damage that befell the third word, by nesting the <supplied> tag within a <damage> tag, as seen below:

*I think its purchase highly desirable as a sample of the style of art of which****<gap>*** *I should be* ***<supplied>****willing****</supplied>*** *to see* ***<damage type=inkblot><supplied>*** *used****</supplied></damage>*** *on any models in the museum.*

Notice that the reason attribute in <supplied> is not necessary if it is nested within a <damage> tag.

4.8 Internal Editorial Remarks

The tags in this section are used to tag text that was altered on the original document, whether by the author or a subsequent reader. Included in this section:

- **<del> Deleted Text**

- **<add> Added Text**

- **<restore> Text Restored to a Previous State**

<del> Deleted Text

- <del> indicates text that was deleted in the original document
- Paragraph tags cannot be inserted within <del>. If the deleted text spans several paragraphs, each paragraph must be marked separately with <del>
- The <del> tag has the attributes type and status available to it
    - The type attribute classifies the deletion. Values for type include:
        - overstrike (crossed-out text)
        - erasure (erased text)
        - bracketed (brackets in the text or margin)
        - subpunction (dots beneath the letters deleted)

    ***Note: You may add to this list, but always remember to *be consistent* and *log your choices*

    - The status attribute identifies faulty deletion. Values for status include:
        - excess start (too much deleted at the start of deletion)
        - excess end (too much deleted at the end of deletion)
        - short start (not enough deleted at the start of deletion)
        - short end (not enough deleted at the end of deletion)
        - unremarkable (deletion not faulty)

    ***Note: You may add to this list, but always remember to *be consistent* and *log your choices*

Examples:

An enamel of the Perricand's style – of rare excellence in colour and process – such a one in fact as very ***<del type=overstrike status=unremarkable>*** *rarely****</del>*** *seldom comes into the market.*

This example shows that the word "rarely" has been crossed-out in the text.

***Note: If the text being deleted cannot be made out with any certainty at all, the <gap> tag should be used in place of <del>.

<add> Added Text

- <add> indicates text was added in the original document

- Paragraph tags cannot be inserted within <add>. If the added text spans several paragraphs, each paragraph must be marked separately with <add>

- The place attribute is used in the <add> tag to specify where the addition is located. Values include

    - supralinear (addition is made above the line)
    - infralinear (addition is made below the line)
    - marginleft (addition is made in the left margin)
    - marginright (addition is made in the right margin)
    - margintop (addition is made in the top margin)
    - marginbot (addition is made in the bottom margin)
    - overleaf (addition is made on the back of the paper)
    - inline (addition is made in a space left by the original author)

***Note: You may add to this list, but always remember to *be consistent* and *log your choices*

Example:

An enamel of the Perricand's style – of rare excellence in colour and process – such a one in fact as very ***<del type=overstrike>*** *rarely* ***</del><add place=supralinear>*** *seldom* ***</add>*** *comes into the market.*

This example shows that the word "rarely" has been crossed-out in the text, with seldom added in its place. Seldom has been added above the line of text.

<restore> Text Restored to a Previous State

- <restore> signifies text that was restored to an earlier state by the author or a later editor. For example, text that was crossed out and then marked to show that it should be kept would be marked with <restore>

- The <restore> tag has the attributes type and desc available to it

    - The type attribute supplies the type of action being cancelled. Values for type include:

        - del
        - add

    ***Note: You may add to this list, but always remember to *be consistent* and *log your choices*

    - The desc attribute is used to supply a prose description of the type of action being undone

Examples:

An enamel of the Perricand's style – of rare excellence in colour and process – such a one in fact as very ***<restore type=del desc='supralinear "stet"'>*** *rarely* ***</restore>*** *seldom comes into the market.*

This example shows that the word "rarely" has been crossed-out, but restored by the writing "stet" above the word.

Alternatively, the <restore> tag could have the <del> tag nested in it to describe the nature of the original deletion.

An enamel of the Perricand's style – of rare excellence in colour and process – such a one in fact as very ***<restore desc='supralinear "stet"'><del type=overstrike>****rarely****</del></restore>*** *comes into the market.*

Notice that the type attribute in <restore> is not necessary if the <del> tag is nested within it.

4.9 External Editorial Remarks

The tags in this section are used for remarks added by the transcriber. Included in this section:

- **<abbr> An Abbreviation**
- **<sic> A Mistake**

- **<orig> Irregular or Antiquated Text**

<abbr> An Abbreviation

- <abbr> contains any form of an abbreviation.
- The <abbr> tag has the attributes expan, cert, and type available to it
    - The expan attribute is used to supply an expansion of the abbreviation.
    - The cert attribute describes how certain you are that the expansion is correct. The value of certainty should be a number between 0 and 100, with 0 meaning that the expansion is definitely wrong and 100 meaning that the expansion is definitely correct.
    - The type attribute describes the type of abbreviation. Values for type include:
        - suspension (provides only the first letter of the word or phrase)
        - contraction (omits some letters in the middle)
        - brevigraph (comprises a special symbol or mark)
        - superscription (includes writing above the line)
        - acronym (comprises the initials of the words of a phrase)
        - title (for a title of address, such as Mr.)
        - organisation (for the name of an organisation)
        - geographic (for a geographic name)

    ***Note: You may add to this list, but always remember to *be consistent* and *log your choices*

Examples:

<persName><abbr type=title expan=Mister cert=100>**Mr.**</abbr>**Digby Wyatt**</persName>

<abbr type=acronym expan="Self Contained Underwater Breathing Apparatus" cert=85>**SCUBA**</abbr>

The first example shows that Mr., within the name Mr. Digby Wyatt, is a title short for Mister, of which the transcriber is 100% certain.

The second example shows that SCUBA is an acronym for "Self Contained Underwater Breathing Apparatus"; however, the transcriber is only 85% certain of this.

<sic> A Mistake

- <sic> signifies text that is known to be contain a mistake
- The <sic> tag has the attributes corr and cert available to it
    - The corr attribute is used to supply the correct version of the text
    - The cert attribute describes how certain you are that the correction is accurate. The value of certainty should be a number between 0 and 100, with 0 meaning that the correction is definitely inaccurate and 100 meaning that the correction is definitely accurate.

Example:

When I have seen Mr. Thompson's photographs from Santiago, Coimbra, & Batalha, I shall be able to ***<sic corr=advise cert=100>****advice****</corr>*** *with more certainty as to his future operations.*

This shows "advice" was used in the original text, where it should have been "advise." The transcriber is 100% certain of the accuracy of this correction.

<orig> Irregular or Antiquated Text

- <orig> is used to contain irregular or antiquated text
- The attribute reg is available to the <orig> tag to contain a regularised version of the text

Example:

how godly a ***<orig reg=deed>****dede****</orig>*** *it is to* ***<orig reg=overthrow>*** *overthrowe****</orig>*** *so wicked a race the world may judge: for my part I* ***<orig reg=think>****thinke****</orig>*** *there* ***<orig reg=cannot>****canot****</orig>*** *be a greater* ***<orig reg=sacrifice>****sacryfice****</orig>*** *to God.*

This example shows how five antiquated spellings were marked, with the regular versions noted in the reg attribute.

4.10 Drawings and Figures

The tags in this section are used to describe pictures and drawings within the Robinson Reports. Included in this section is:

- **<figure> A Figure**
- **<head> A Heading**
- **<figDesc> A Prose Description of a Figure**

<figure> A Figure

- <figure> is used to mark where a picture of some sort existed in the original drawing
- The <head> and <figDesc> elements described below can be used as subelements of the <figure> tag to provide more information on the figure

See <figDesc> for a complete example

***Note: Figures can actually be inserted using the entity attribute of <figure>, but this is a very complicated process. We recommend you only provide descriptions and locations of figures, so they can be inserted at a later date.

<head> A Heading

- While <head> can be used as any kind of heading, here it is used specifically as a subelement of <figure> to represent the caption of a figure

See <figDesc> for a complete example

<figDesc> A Prose Description of a Figure

- <figDesc> is a subelement of <figure> that is used to supply a prose description of that figure

- This description can be used in place of the actual image if it is not available or in conjunction with it if the image is available

Example:

<figure>
<head>*Figure One: The View from the Bridge***</head>***
<figDesc>*A Whistleresque view showing four or five sailing boats in the foreground, and a series of buoys strung out between them.***</figDesc>***
</figure>

This example shows how a figure would be inserted, with a caption and a prose description.

4.11 Annotations, Comments, and Interpretations

The tag in this section is used to tag notes on the original document or comments from the transcriber. Included in this section:

- **<note> A Note**

<note> A Note

- <note> could contain a minute or annotation on the original document or a comment or interpretational note from you

- The <note> tag has the attributes type, resp, place and anchored available to it

    - The type attribute specifies the type of note being entered. Legal values for type include:

        - minute (notes written on the original by previous readers)
        - comment (your comments on the report)

      The *minute* type often has a signature of the writer at its end. However, there is no way to denote a signature within a <note> tag. Instead, simply tag the name with <persName>, described in section 4.3 *Names*. The resp attribute described below can then be used to assign responsibility to that person.

      More information on where and how to insert the different types of <note> tags can be found in chapter 3 *Templates*.

    - The resp attribute gives the name of the notes author. If you wrote the note, the resp attribute should be set to *transcr*; otherwise, it should be set to the name of the author. If the author is unknown or illegible, set the value of resp to *unknown*.

    - The place attribute specifies where on the report the note appears. Valid values include:

        - head (appears at the head of the page)
        - foot (appears at the foor of the page)
        - left (appears in the left margin)
        - right (appears in the right margin)
        - interlinear (appears within the text)
        - back (appears on the back of the report)
        - footnote (appears in a footnote)

    - The anchored attribute indicates whether the transcription shows the exact location of the note on the original. anchored has two legal values: *yes* and *no*. For the Robinson Reports, the value of anchored should always be *no*. The default value for anchored is *yes*; therefore, every time a note is added, the anchored tag must be changed from *yes* to *no*.

Example:

<note anchored=no type=annotation resp="Richard Redgrave" place=foot>*The price of such works as this is determined by their variety – I concur with Mr. Wyatt – but think it hardly desirable as a purchase for our museum.****<lb><persName reg="Redgrave, Richard">****Rick R.****</PersName></note>***

<note anchored=no type=comment resp=transcr>*The edges of this report are very tattered and faded.****</note>***

<note anchored=no type=interp resp=transcr>*The third paragraph, at the end of sentence two. Robinson is referring to Thurston Thompson's journey to Santiago to take photographs.****</note>***

These examples show one of each type of note. Notice how the writer's signature is handled using the <persName> tag in the first example. Also note the initial sentence of the third example, describing the location of the reference.

4.12 Lists

The tags in this section are used to create lists. Lists should be used whenever any type of listing of items, with any number of columns, is included within one of the Robinson Reports. A template is available for reports that are entirely made up of a list; however, the template employs the tags and methods described here, so a familiarity with the information presented here will be beneficial when using it. Included in this section:

- **<list> A List**
- **<item> A List Item**
- **<label> A Column Label**

These items will be described separately, followed by instructions and examples for constructing a list.

<list> A List

- <list> is used to denote a list of any type, having any number of columns and rows, and with optional headers on the columns

- The tags <item> and <label> are sub-elements of <list>, and must be used within it

- The <list> tag has the attributes type and n available to it

    - The type attribute is used to describe the type of list being entered. Acceptable values include:

        - ordered (list is numbered or lettered in some way)
        - bulleted (list is bulleted)
        - simple (list is neither ordered nor bulleted)
        - otable (ordered list with more than one column)
        - btable (bulleted list with more than one column)
        - stable (simple list with more than one column)

    ***Note: You may add to this list, but always remember to *be consistent* and *log your choices*

    - The n attribute holds the number of columns in the list

---

<item> A List Item

- <item> contains a single item in the list
- If a list is only one column, the <item> tag represents the entire row; if a list has more than one column, the <item> tag represents the text in one column of one row
- <item> is a subelement of <list> and can only be used within it
- The n attribute in <item> indicates the number or letter of the list item in an ordered list

---

<label> A Column Label

- <label> signifies the label of a column in a list
- If a list has no labels, this tag is not necessary; if a list has a label on some columns, but not all columns, empty <label> tags must be included for the unlabeled columns
- <label> is a sub-element of <list> and can only be used within it

---

Creating a List without Labels

1. Insert a <list> tag where you want the list to appear. Set the n attribute to the number of columns in the list. Set the type attribute to the type of list you are making
2. Insert an <item> tag within the <list>
3. If there is a second column, insert another <item> tag after first
4. Continue inserting item tags for each additional column. A list with four columns should look something like this:

   <list n=4 type=otable>
   <item></item> <item></item> <item></item> <item></item>
   </list>

   This represents one row of the list.

5. Copy everything within the <list> tag by highlighting it with the mouse and holding **Ctrl** and pressing **C**. Do not copy the actual <list> tag. All of the rows appear within one <list> tag, so it does not need to be duplicated.
6. For each additional row in the list, paste in a copy of the first row.
7. Now, if you are creating an ordered list, set the n attribute of the first <item> tag in each row to the number or letter of the row. For a list with four columns and four rows, you should now have something that looks like this:

   <list n=4 type=otable>

   <item n=1></item> <item></item> <item></item> <item></item>
   <item n=2></item> <item></item> <item></item> <item></item>
   <item n=3></item> <item></item> <item></item> <item></item>
   <item n=4></item> <item></item> <item></item> <item></item>

   </list>

10. Now, to complete the list, fill the information from the list into the item tags.

Example:

The list:

1. An Italian Vase	£12.0.0
2. A Gate of a Cathedral	£117.5.2
3. A Marble Chessboard	£7.0.7
4. Some Lace	£0.9.0
5. A Small Statuette	£11.0.0
6. A Silver Pocket Watch	£27.10.0
7. A Rifle	£43.0.0
8. A Monk's Cowl	£0.4.1

Should be encoded as follows:

<list n=2 type=otable>
<item n=1>*An Italian Vase****</item><item>****£12.0.0****</item>*

<item n=2>*The Gate of a Cathedral****</item><item>****£117.5.2****</item>*

<item n=3>*A Marble Chessboard****</item><item>****£7.0.7****</item>*

<item n=4>*Some Lace****</item><item>****£0.9.0****</item>*

<item n=5>*A Small Statuette****</item><item>****£11.0.0****</item>*

<item n=6>*A Silver Pocket Watch****</item><item>****£27.10.0****</item>*

<item n=7>*A Rifle****</item><item>****£43.0.0****</item>*

<item n=8>*A Monk's Cowl****</item><item>****£0.4.1****</item>***
</list>

Creating a List with Labels

8. Insert a <list> tag where you want the list to appear. Set the n attribute to the number of columns in the list. Set the type attribute to the type of list you are making
9. Insert a <label> tag inside the <list> tag. An <item> tag should be automatically inserted beside the <label> tag.
10. Type the label of the first column in the <label> tag.
11. If there is a second column, insert another <label> tag after the <item> tag.
12. Type the label of the second column in the second <label> tag.
13. Repeat steps 4 and 5 for each additional column. At this point, you should have a <label> and <item> tag for each column. A list with four columns should look something like this:

    <list n=4 type=otable>
    <label>Label 1**</label><item></item><label>**Label 2**</label><item></item>**
    <label>Label 3**</label><item></item><label>**Label 4**</label><item></item>**
    </list>

    This represents one row of the list.

14. Copy everything within the list tag by highlighting it with the mouse and pressing Ctrl-C.
15. For each additional row in the list, paste in a copy of the first row.
16. Now, if you are creating an ordered list, set the n attribute of the first <item> tag in each row to the number or letter of the row. For a list with four columns and four rows, you should now have something that looks like this:

    <list n=4 type=otable>
    <label>Label 1**</label><item n=1></item><label>**Label 2**</label><item></item>**
    <label>Label 3**</label><item></item><label>**Label 4**</label><item></item>**

    <label>Label 1**</label><item n=2></item><label>**Label 2**</label><item></item>**
    <label>Label 3**</label><item></item><label>**Label 4**</label><item></item>**

    <label>Label 1**</label><item n=3></item><label>**Label 2**</label><item></item>**
    <label>Label 3**</label><item></item><label>**Label 4**</label><item></item>**

    <label>Label 1**</label><item n=4></item><label>**Label 2**</label><item></item>**
    <label>Label 3**</label><item></item><label>**Label 4**</label><item></item>**
    </list>

10. Now, to complete the list, fill the information from the list into the item tags.

Example:

The list:

Object	Price
1. An Italian Vase	£12.0.0
2. A Gate of a Cathedral	£117.5.2
3. A Marble Chessboard	£7.0.7
4. Some Lace	£0.9.0
5. A Small Statuette	£11.0.0
6. A Silver Pocket Watch	£27.10.0
7. A Rifle	£43.0.0
8. A Monk's Cowl	£0.4.1

Should be encoded as follows:

<list n=2 type=otable>
<label>Object***</label><item n=1>***An Italian Vase***</item>***
<label>Price***</label><item>***£12.0.0***</item>***

<label>Object***</label><item n=2>***The Gate of a Cathedral***</item>***
<label>Price***</label><item>***£117.5.2***</item>***

<label>Object***</label><item n=3>***A Marble Chessboard***</item>***
<label>Price***</label><item>***£7.0.7***</item>***

<label>Object***</label><item n=4>***Some Lace***</item>***
<label>Price***</label><item>***£0.9.0***</item>***

<label>Object***</label><item n=5>***A Small Statuette***</item>***
<label>Price***</label><item>***£11.0.0***</item>***

<label>Object***</label><item n=6>***A Silver Pocket Watch***</item>***
<label>Price***</label><item>***£27.10.0***</item>***

<label>Object***</label><item n=7>***A Rifle***</item>***
<label>Price***</label><item>***£43.0.0***</item>***

<label>Object***</label><item n=8>***A Monk's Cowl***</item>***
<label>Price***</label><item>***£0.4.1***</item>***
</list>

4.13 An Index to the Tags

<abbr> An Abbreviation 94
<add> Added Text 92
<collectionName> Name of a Collection 74
<cost> Cost of a Purchase 78
<damage> Damaged Text 87
<date> A Date 76
<dateRange> A Range of Dates 77
<del> Deleted Text 91
<distinct> Distinct Words or Phrases 82
<emph> Emphasized Text 82
<eventName> Name of an Event 75
<figDesc> A Prose Description of a Figure 97
<figure> A Figure 96
<foreign> Foreign Words or Phrases 83
<gap> Unreadable Text 88
<handShift> Change in Hand 85
<head> A Heading 97
<hi> Graphically Distinct Text 81
<item> A List Item 100
<label> A Column Label 100
<lb> Line Break 69
<list> A List 99
<materialName> Name of a Material 74
<measure> A Quantity 79
<mentioned> Mentioned Words or Phrases 84
<name type=site> Name of Specific Site 75
<note> A Note 98
<num> A Number 79
<objectIdentifier> Museum Object Number 78
<objectName> Name of an Object 73
<orgName> Name of an Organisation 72
<orig> Irregular or Antiquated Text 95
<p> Paragraph 68
<pb> Page Break 69
<persName> Name of a Person 71
<placeName> Name of a Place 72
<quote> A Quotation 81
<restore> Text Restored to a Previous State 92
<sic> A Mistake 95
<soCalled> Unusual Words or Phrases 84
<space> Blank Space 85
<supplied> Text Supplied by the Transcriber 90
<unclear> Unclear Text 87
Abbreviation 94
Added Text 92
ana attribute -- Analysis and Interpretations of Elements 67
Analysis and Interpretations of Elements 67
Annotations 98
Antiquated or Irregular Text 95
Caption 97
Comments 98
corresp attribute -- Correspondence between Elements 67
Correspondence between Elements 67
Cost of a Purchase 78
Damaged Text 87
Date
Date 76
Range 77
Deleted Text 91
Description of a Figure 97
Distinct Words or Phrases 82
Drawing 96
Element Identifier Attribute 64
Emphasized Text 82
Figure 96
Foreign Words or Phrases 65, 83
Graphically Distinct Text 66, 81
Hand Change 85
Heading 97
id attribute -- Element Identifier 64

Illegible Text
Damaged Text 87
Text Supplied by the Transcriber 90
Unclear Text 87
Unreadable Text 88
Interpretations 98
Irregular or Antiquated Text 95
lang attribute -- Foreign Words or Phrases 65
Line Break 69
Lists 99
Column Labels 100
Creating a List with Labels 103
Creating a List without Labels 101
List Items 100
Mentioned Words or Phrases 84
Minutes 98
Mistakes 95
Museum Object Number 78
n attribute -- Name or Number 65
Name
Collection 74
Event 75
Material 74
Object 73
Organisation 72
Name
Person 71
Place 72
Site 75
Name or Number Attribute 65
Notes 98
Number 79
Page Break 69
Paragraph 68
Picture 96
Quantity 79
Quotation 81
Unusual Words or Phrases 84
Quotation Marks
Graphically Distinct Text 81
Quotation 81
rend attribute -- Graphically Distinct Text 66
Restored Text 92
Space 85
Supplied Text 90
Unclear Text 87
Unreadable Text 88
Unusual Words or Phrases 84

5 Shortcuts and QuickWords

Several shortcuts specifically tailored to the Robinson Reports have been defined in WordPerfect. Shortcuts to seven phrases have been defined in the form of QuickWords. If you type an abbreviated form of these phrases, WordPerfect will automatically expand them when a space is entered. For example, if you type *+JR* followed by a space, it will be expanded to *J.C. Robinson*. For a QuickWord to expand, **you must enter a space or a carriage return after typing it**. For a complete list of QuickWords, see Table 3.

Table 3 - QuickWords defined in WordPerfect

Common Phrase	**Abbreviated Form**
J.C. Robinson	+JR
Mr. Digby Wyatt	+DW
South Kensington Museum	+SKM
Kensington Museum	+KM
Science and Art Department	+SAD
Your obedient servant,	+YOS
Your most obedient servant,	+YMOS

Notice that all of the QuickWords start with a "+"; this is done to delineate them from actual text that may come up in the transcriptions. For instance, if "JR" was the abbreviated form for "J.C. Robinson", the editor would expand the text "JR" at times when it wasn't meant to.

In addition to the QuickWords, the dollar sign key on the keyboard has been altered to enter a pound sign. Because of this, you will not have to use the insert symbol function every time you wish to insert a pound sign.

***Note: Due to shortcomings in the WordPerfect 9 editor, these shortcuts will not work in prompts or other dialog boxes. If these phrases appear fields that you are prompted for, you must type them out.

6 Appendix A - Reference Numbers for the Boxes

Table 4 - Reference Numbers for the Art Referee's Reports, 1863-1886								
J.C. Robinson's reports					**Art Referees' reports**			
Sticker#	Reference#	Vol#	Part#		Sticker#	Reference#	Vol#	Part#
1A	MA/3/1	I	1		15A	MA/3/27	I	1
1B	MA/3/2	I	2		15B	MA/3/28	I	2
2A	MA/3/3	I	3		16A	MA/3/29	I	3
2B	MA/3/4	I	4		16B	MA/3/30	I	4
3A	MA/3/5	I	5		17A	MA/3/31	II	1
3B	MA/3/6	I	6		17B	MA/3/32	II	2
4A	MA/3/7	II	1		18A	MA/3/33	II	3
4B	MA/3/8	II	2		18B	MA/3/34	II	4
5A	MA/3/9	II	3		19	MA/3/35	III	1
5B	MA/3/10	II	4		20	MA/3/36	III	2
6A	MA/3/11	II	5		21A	MA/3/37	III	3
6B	MA/3/12	III	1		21B	MA/3/38	III	4
7A	MA/3/13	III	2		22A	MA/3/39	IV	1
7B	MA/3/14	III	3		22B	MA/3/40	IV	2
8A	MA/3/15	III	4		23A	MA/3/41	IV	3
8B	MA/3/16	IV	1		23B	MA/3/42	IV	4
9	MA/3/17	IV	2		24A	MA/3/43	V	1
10	MA/3/18	IV	3		24B	MA/3/44	V	2
11A	MA/3/19	IV	4		25A	MA/3/45	VI	1
11B	MA/3/20	V	1		25B	MA/3/46	VI	2
12A	MA/3/21	V	2		26A	MA/3/47	VI	3
12B	MA/3/22	V	3		26B	MA/3/48	VII	1
13A	MA/3/23	V	4		27A	MA/3/49	VII	2
13B	MA/3/24	VI	1		27B	MA/3/50	VII	3
14A	MA/3/25	VI	2		28A	MA/3/51	VIII	1
14B	MA/3/26	VI	3					

7 Appendix B – Sample Letters

3

28. 2

April 28th 1863.

I recommend for purchase of Mr G. Eastwood a casket in brass or "latten", of the 15th century, engraved with animals, foliated ornamentation and inscriptions in "Gothic" or Church text characters. The Museum does not possess any specimen of this class, which is a known and familiar one.

The price is £5-0-0.

[illegible] as to be sent before 6 May

done 4.5.63.

purchased sanctioned: 6 May 63

H.C.

~~[illegible]~~

And also of Mr Whitehead a statuette of a child or "Amorino" in carved pear or boxwood, on a pedestal of Boule work.

Price £60.

The statuette is of German 16th century origin, the pedestal probably made for it circa 1750. (?)

postponed

It is a fine & characteristic specimen of its kind, the work of an eminent sculptor or woodcarver - Mr Whitehead obtained it for £50 from Mr Brett during its continuance in the recent "Loan Exhibition". See notice "Loan Exhibition Catalogue" No 6740.

J C Robinson

return: 4/ May 63

Addendum

MA-3-1-11

Larghezza totale metri 1 e 70 centimetri

Group No 2

35.

altezza, centimetri 98

Basic Letter

MA-3-1-3

28 Apl. 63

Recommends the purchase for £5 of a Casket from Mr. Eastwood and of a Statuette for £60 from Mr. Whitehead.

Mr Cole
30 Apr
4. May

Eastwood
letter drafted accordingly

[illegible]

May 5 Written
[illegible] 4.5.63

SCIENCE AND ART
No 8408
30 APR
1863

Form Report

MA-3-46-97-1

97/1

July 1871. No. 652.

Noted

Science and Art Department,

South Kensington Museum,

April 9 1875.

Madam,

I have to request you to visit the Museum and report upon the objects enumerated below.

The documents connected therewith will be submitted to you upon application to the officer in charge of the Art Museum Division.

A separate report upon each case referred for your opinion is requested. The reports should be written on the accompanying forms or on foolscap paper, half margin.

I am, Madam,

Your obedient Servant,

To Mrs F. Palliser

Nauenn Macleod

GW.

SUBJECTS FOR REFERENCES.

1. A collection of ~~Lace~~ Embroidery £925–

The altar fronts 9 & 10 are desirable, 9 for its oriental style, 10 for its harmonious colouring but the prices are too preposterous to entertain. All the pieces sent in bear the appearance of having been transferred upon fresh velvet, the gold fringes are new, & there is a great deal of patching & piecing for sale. It may be a question whether for the artistic merit of the pieces as regards colour & arrangement it is worth while to purchase specimens for the S.K.M. of which so little remains of their identity

Noted 26 ap. Rt.

Fanny Palliser

List of Objects

MA-3-11-349

List of Objects constituting the Purchases from the Campana Collection as Registered in the Art Division.

South Kensington Museum

7559	1861	Statue in Marble Virgin Kneeling by Matteo Civitale H 3 ft 5 in W 16 in	120	
7560	"	Statue in Marble Cupid resting on one Knee his left arm uplifted a quiver &c at his side By Michael Angelo Believed to be the one executed for Jacopo Galli in 1497 Base 19½ x 18 in H 3 ft 5½ in	1000	
7561	Jany	Unfinished Statuette in carved marble St Sebastian by Michael Angelo Buonarroti Base 12 in x 11 in H 3 feet	250	
562	"	Alto relievo in marble partly gilt Half length figure of the Virgin with the Infant Saviour ascribed to Benedda Rovezzano H 1 ft 10½ in W 1 ft 2½ in	25	
7563	"	Alto relievo in marble The annunciation at the sides are respectively figures of the Virgin and the announcing angel between them under a Canopy supported by two twisted columns is a statuette on a smaller scale of Our Saviour H 2 ft 5 in W 4 ft 2 in	50	0

1445

7564	1861 Jany.	Relievo in marble. Virgin seated with the Infant Jesus. flanked on either side by an angel. Tuscan sculpture circa 1400 H 2ft 3½ in. W 2ft 8 in	30	0	0
7565	"	Bas relief in marble. The Virgin (or a female saint) accompanied by an angel holding a lily-branch, giving money to two female children. Tuscan sculpture. circa 1440 N. 2ft 2½ in. W 2ft. 5½ in	40	0	0
7566	"	Fragmentary alto relievo in marble from a mural sepulchral monument. Statuette of an angel. School of the Pisani circa 1340 N. 2ft 1 in W. 1ft. 5 in	15	6	0
7567	"	Similar alto relievo in marble. Statuette of an angel. attributed to the Pisani N. 2ft 1in W 1ft.			
7568	"	Tabernacle in black stone, carved in bas relief with ornament of renaissance design. Above is a lunette shape compartment containing a representation of the Holy Spirit as a dove, surmounted by a honeysuckle ornament. The whole supported on a bracket composed of two cornucopiae enclosing a cherub Period of Desiderio da Settignano. circa 1460-80 N. 5ft W 2ft 1 in	20		

Minute Paper

MA-3-16-47

47 38

June 1865. No. 381.

MINUTE PAPER.

Science and Art Department.

No. of Paper.

SCIENCE AND ART No 18 28 AUG 1865

No. of Enclosures.

NAMES.	DATES.	MINUTES.
Mr Cole	20/9/65	This report has been retained in error by Mr Black during my absence. It was referred to him for the preparation of the label for the picture as a loan, as I thought it very desirable that Mr Robinson's suggestion should be strictly followed. I told Mr Thompson that it was not desirable to place the work in a very prominent place as in my opinion the less it is seen, until received by purchase, the better. [illegible]
Mr Cole	22 Sept	1 See Minute 30/11.65 not to be informed till further orders H C
		2 Accountant 1 Dec enter as [illegible] H C Entered [illegible]
		Return 4.12.65 Keep

Minute Paper

MA-3-46-71

71

December, 1874.

S. K. M. Form No. 44.

MINUTE PAPER.

South Kensington Museum.

Registered No.

Registry

Correspondence 25 Jan 75

Corn. 26.1.75

Report enclosed.

D. 27/1

Please to make another copy and draft letter to Mr Edwards enclosing report and requesting him to inform the Museum should he be disposed to part with his ~~[illegible]~~ Miniatures and the price in order that the question of acquiring the same might be brought under ~~the~~ consideration. The report upon the Miniatures has been carefully prepared for labelling purposes & probably will be of interest. P.O.

25 Jan 75

ProfRefPaper

MA-3-27-49

U.R.R. 49 No. 3

March 1868. No. 653.

SCIENCE AND ART DEPARTMENT OF THE COMMITTEE OF COUNCIL ON EDUCATION, SOUTH KENSINGTON MUSEUM.

SCIENCE AND ART No. 17583 6 MAY 1868

PROFESSIONAL REFERENCE PAPER.

1. Subject of Reference. on A Bronze Medallion "Marriage of Henry IV" from Mr G. G. Brooks
2. Price. £1.0.0 (Registered No. of document 15040)
3. Reference to similar objects in the Museum.

Year of acquisition . Registered No. . Price. £

REPORT.

read 18d 8 May 68

This I believe to be a poor galvano-plastic reproduction of the reverse of a very fine medallion by Duprét of which the Museum already possesses an admirable original.

I do not recommend the purchase

M. Digby Wyatt

M A
6.5
14.5 Noted J.B.R.
GW 7.5.68

See 15040
Article returned 19/5

19/5

ProfRefPaper(Rec From)

MA-3-46-127

a. R. R.
Entd. AMB. 3/8/75

127

January, 1870. No. 653.

SCIENCE AND ART DEPARTMENT
OF THE COMMITTEE OF COUNCIL ON EDUCATION,
SOUTH KENSINGTON.

SOUTH KENSINGTON MUSEUM No. 3129 10 JUN 1875

PROFESSIONAL REFERENCE PAPER.

1. Subject of Reference on a collection of Lace
2. Received from Mr. Chick
3. Price Various prices. (Registered No. of document)
4. Reference to similar objects in the Museum

Year of acquisition Registered No. Price £

REPORT.

Date and Country.

I have examined two parcels of Mr Chick's lace, there is nothing in them that is new & of which examples are not already in the Museum Collection.

Nos 428 & 585 are two pretty border laces of uncommon pattern that might be taken.

649 Venetian shawl is of very pretty design but would appear to be made up in strips to meet its present form.

We have two much of this class of lace.

Mrs. F. Pallison

Noted. 9 June RL

seen Return with thanks (on Aguada Fd) R. A. Thompson Actuary

A. M. 8. 6. 75

10/6 ? Done with

Discounted 5/7/75

F & T 500 6—72

Telegraph

MA-3-16-53-3

(277 W.)
(Forwarded Form.)

53/30 No. of Message 50

THE ELECTRIC AND INTERNATIONAL TELEGRAPH COMPANY, _____ Station.

Prefix ____ Code Time ____ WORDS TO BE SIGNALLED ____

Received ____ m
Finished ____ m

Date ____ 186
Sent to ____ Station
by me ____ Clerk.

Message ..... " "
Repeating ..... " "
Reply ..... " "
Porterage ..... " "
To be paid out ..... " "
Total ..... " "

(D Q—M M) (Address) (M M—P Q)

Counter Clerks' Initials .....

FROM

Name and Address of the Sender of the Message.

Secretary
South Kensington Museum

TO

Name and Final Address of the Person to whom the Message is to be delivered.

I. C. Robinson
British Embassy - Lisbon
D Q

Five hundred sent to Vanzeller of Lisbon twenty third September - letter Post Office. Brucciani ordered to meet you on his way to Santiago - telegraph place and date for meeting.

Please to Telegraph the above UNINSURED MESSAGE according to the conditions endorsed hereon; and forward such Message from the Terminal Station of the Company at _____ by _____ to the above Address, subject to the endorsed conditions, for which latter purpose I have deposited _____

The latter part of this request only applies to messages requiring a deposit.

You are requested before signing to read the Conditions of the Contract on the back.

Signature and Address of Sender.

The Company will not be answerable for Errors caused by indistinct writing.

13 Appendix G – Guidelines for Completing the System

Guidelines for Completing the System

Written By

Chris Holt
Micah Kiffer
Keith Peterson

Completed as a part of an Interactive Qualifying Project for Worcester Polytechnic Institute, in conjunction with the National Art Library at the Victoria and Albert Museum in London

Table of Contents

1 INTRODUCTION 1
2 INTERPRETING THE XML DOCUMENTS 2
2.1 The Tags *2*
2.2 The teiHeader *3*
2.3 The Back Section *6*
2.4 The Main Section *7*
2.5 The Notes Sections *7*
3 COMPLETION OF THE SYSTEM 9
3.1 Corpus Files *9*
3.2 Further Tagging of the Transcriptions *9*
3.3 XSL Style Sheets *10*
3.4 Implement the Search System *11*

1 Introduction

This manual will explain how to interpret the overall structure of the XML documents created for the transcriptions as well as give recommendations on what needs to be done to achieve the final goals for the Robinson Reports. The vision the National Art Library (NAL) has for the Robinson Reports is to produce electronic versions of the Robinson Reports so that the information they contain will be more accessible to scholars, and to develop an electronic search system that will allow scholars to search the reports using specific search criteria. This manual explains what we have done to facilitate the project and what needs to be done for its completion. The manual consists of two chapters:

- *Interpretation of the XML Documents*
- *Completion of the System*

Each of the XML documents follows a certain structure. The *Interpretation of the XML Documents* chapter will explain the basic layout of each document and how the tags used to mark up the document should be interpreted. This information will be helpful when interpreting the XML document for display or search purposes.

After Professor Fontanella has completed the transcriptions, more work will need to be done in order to accomplish the goals that the NAL has defined for the Robinson Reports. The *Completion of the System* chapter will explain the recommended steps for completing the system.

For more information on the work done with the Robinson Reports, see the full project report, *Transcription and Cataloguing of the Robinson Reports*, and the *Transcription Manual.*

2 Interpreting the XML Documents

All of the documents are divided into sections. With only a few exceptions, every report has a header section, back section, main section, minutes and annotations section, and a comments section. The only documents that do not follow this structure are telegraphs and preliminary reports, which do not have a back section, and minute paper, which does not have a minutes and annotations section. This chapter will explain each of the sections in detail. First, however, it will describe how to interpret the tags that are used to mark up the documents.

2.1 The Tags

The definitions and uses of all TEI tags are available in the *TEI Guidelines for Electronic Text Encoding and Interchange (P3)*; however, some tags were interpreted in a unique manner for the Robinson Reports, while others employed unique attribute values. The interpretations of the tags and all of the unique attribute values are given in detail in the *Transcription Manual* included with this report. The tags that had custom interpretations and attribute values applied to them include <note>, <list>, <label>, <item>, <quote>, <hi>, <persName>, <name>, <measure>, <damage>, <gap>, <supplied>, <del>, <add>, and <restore>. The six CIMI tags that were used all have special interpretations and uses as well. These include <objectName>, <materialName>, <eventName>, <collectionName>, <objectIdentifier>, and <cost>. It is highly recommended, however, that the entire *Transcription Manual* be reviewed.

In addition to the *Transcription Manual*, Professor Fontanella was asked to keep track of any additional attribute values that he defined. If he did define his own attribute

values, that documentation should be available as well. Since the transcriptions were not completed at the time of printing, this document cannot disclose the attributes from the *Transcription Manual* that were actually used or the attribute values that Fontanella defined.

There is one very important tag that the *Transcription Manual* does not describe: the tag. The tag is used to divide the document into its various components (back, main, minutes and annotations, and comments). The n attribute in the tag is set to the name of the section. Note that *Professional Reference Paper* and *Professional Reference Paper(Rec From)* use the <div0> tag in place of the tag; this is done because they are further subdivided into columns using the <div1> tag.

***Note: many tags in the documents will be empty. The templates used by Fontanella in the transcriptions often had tags defined that did not appear in every letter. Fontanella was instructed to leave these tags empty.

2.2 The teiHeader

The teiHeader, found at the beginning of each document, is not documented in the *Transcription* Manual either. The teiHeader contains bibliographic information about the electronic and copy texts, as well as the document type. The teiHeader, seen in Figure 1, is constructed exactly the same way in each of the transcriptions. For purposes of interpretation, the teiHeader for the Robinson Reports has been split into four sections: the title statement, the publication statement, a description of the source material, and the document type. The first three sections are a part of the <fileDesc> element and the fourth is a part of the <profileDesc> element. For more information on these elements and their functions, see the TEI documentation.

The first section, denoted section A in Figure 1, describes the title of the report and a statement of responsibility for the transcriptions. The title gives the name of the collection, which can be either the *J.C. Robinson's reports* or the *Art Referees' Reports*, as well as the particular report's identifying numbers. Also contained within the title statement is the <respStmt> tag, which signifies that Lee Fontanella was responsible for the encoding of the documents and that the nature of his responsibility was transcription.

Figure 1 – The TEI Header

Section B is the publication statement. The publication statement for the Robinson Reports states that the National Art Library at the Victoria and Albert Museum is the authority responsible for making the transcriptions available.

The bibliographic information about the copy text is held in the <sourceDesc> tag, depicted in section C of Figure 1. This section holds all of the information required for a complete bibliographic citation of the report. It includes the name of the NAL and V&A within <orgName> tags, the title of the collection within a <title> tag, and all of the identifying numbers within <biblScope> tags. The <biblScope> tags are differentiated by the type attribute; the values for the <biblScope> tag are, in order of appearance, *volume*, *part*, *reference* (for the box number), *item* (for the individual report number), and *superceded reference* (for the registered paper number, which is an additional cataloguing number written on the Science and Art Department stamp that appears on many of the letters). The *superceded reference* will not be included in every letter; however, because of the use of templates, the tag will always be present, even if it is empty. Notice that the bibliographic information is not formatted in any way and contains no punctuation; this will allow the information to be drawn out and formatted in any desirable manner. It will also allow individual pieces of information to be used without using the entire citation.

The final section of the teiHeader, section D in Figure 1, is used to describe the document type. The <textClass> tag, within the <profileDesc>, was used for this purpose. These tags were used to specify which of the nine document types the text belongs to, which will allow the correct formatting to be applied. All of the documents of each type are structured exactly the same; by reading the document type in the

<textClass> tag, it will be obvious which type of formatting must be applied to the document. There are nine different document types that can be found in the <profileDesc>:

- *Addendum*
- *Basic Letter*
- *Form Report*
- *List of Objects*
- *Minute Paper*
- *Preliminary Reports*
- *Professional Reference Paper*
- *Professional Reference Paper(Rec From)*
- *Telegraph*

For examples of each of these document types, see the *Transcription Manual* included with this document.

2.3 The Back Section

The first part of the actual transcribed text in the report is the back section. The back section provides any information that was originally on the back of the letter. This information is provided at the beginning of the transcription because it was generally written on the outside of the letter, after it had been folded, as an introduction. The back section exists in all document types except for *telegraphs* and *preliminary reports*; note, however, that the back section will very often be left completely empty because of the lack of letters that actually had text on the back. When this is done, the tags will still be present but will have no content.

The back section is designated by a tag with its n attribute set to *Back*. The back section also contains an opener, with a date tag in it, and the body of the text. Any of these tags could be left blank, if they are not applicable, or additional tags could be

present; Fontanella may choose to add tags to the back section that were not included in the templates and are not documented here.

***Note: Professional Reference Paper and Professional Reference Paper(Rec From) use the <div0> tag, not the tag.

2.4 The Main Section

The next section of the transcriptions is the main section of the document; this section of the document is unique to each document type. The main section is delineated by a tag with no name given in the n attribute. It includes information from the body of the report, including openings, closings, special fields, and the text of the report itself.

Keep in mind that some of the printed forms had headers on them that were standard to all forms of that type. These headers were often left out of the transcriptions and will need to be supplied.

***Note: Professional Reference Paper and Professional Reference Paper(Rec From) use the <div0> tag, not the tag. The <div1> tag is used within the main section to further divide the text into two columns.

***Note: The main section of Minute Paper consists entirely of <note> tags. See section 2.5 *The Notes Sections* for information on dealing with minute paper.

2.5 The Notes Sections

The final two sections of the transcriptions are for annotations and comments. The first of these sections has its n attribute set to *minutes and annotations*; the second has its n attribute set to *comments*. Both of these sections consist of <note> tags. For

instructions on how to interpret the <note> tags and their various attributes, see the TEI Documentation and the *Transcription Manual* included with this document.

In addition to appearing in the last two sections of each document, <note> tags are sometimes found in the main section of the transcription. These <note> tags were inserted as footnotes and are related to the text directly preceding them in the document. These <note> tags should have their place attribute set to *footnote* and can be either comments or minutes, as specified by the type attribute.

Note that minute paper does not have a minutes and annotations section. Minute Paper is made up entirely of minutes and annotations, so they are held within the main section of the document. <note> tags in this section are treated exactly the same as <note> tags in other sections, except that they can have *body* as the value for their place attribute.

3 Completion of the System

This chapter discusses four tasks that are recommended for completing the final system: create corpus files, further tagging of the reports, the design of XSL style sheets for each of the document types, and implement a searching system.

3.1 Corpus Files

After Professor Fontanella has completed his transcriptions, each transcription will be stored in its own separate file. The first task that is recommended is to incorporate all of these files into one or more files. This will depend upon how the NAL wants to group the transcriptions. Grouping the transcriptions into a smaller amount of files will help with searching capabilities as well as file management. These transcriptions will have to be grouped together using a tag known as the `<teiCorpus.2>` tag. The corpus tag allows several TEI documents to be defined within it. To combine the transcriptions, copy and paste all of the finished TEI documents into the `<teiCorpus.2>` tag. The corpus tag contains its own TEI header so that the corpus can contain its own bibliographic information.

3.2 Further Tagging of the Transcriptions

Further tagging needs to be done before the system can reach its full potential. Professor Fontanella will be limited in time and will not be able to fully tag all of the reports. A person familiar with the XML and TEI will need to read through the transcriptions and any additional tags that they feel are necessary. For instance, additional tags may want to be included around the names of people. Professor

Fontanella will have only a limited amount of time, so he will be tagging the whole name as one. To provide further searching capabilities tags may want to be provided separating the surname and the forename. Also, attributes of several tags may want to be expounded on. For instance, several of the tags have an attribute called reg that allows the transcriber to regularise names of people and places. This will also help with the searching capabilities of the system because users will probably fill in the full name rather than an abbreviated version.

Also, the capability of inserting cross-references provided by TEI should be incorporated into the transcriptions. Cross-references can be made between related letters or between corresponding elements. When Professor Fontanella records a relationship between two or more reports, <link> tags can be used to join them. Furthermore, if two elements are related, such as an <objectName> and <cost> element, they can be marked as corresponding to each other using the id and corresp attributes. For more information on how to use referencing see the TEI documentation on the website <http://etext.virginia.edu/TEI.html>.

3.3 XSL Style Sheets

Since XML documents contain no formatting, XSL style sheets should be developed for each of the document types. These style sheets are used to specify how the document is to be displayed. The development of the style sheets will be primarily up to the designer of the system.

3.4 Implement the Search System

It is recommended that a system be implemented that is able to use specific search criteria entered by a user to extract only the transcriptions that meet the criteria. There are two ways that this system could be implemented. The system could be made so it that it searches the actual transcription documents looking for relevant information. Another possibility is to implement a program that parses all the transcription documents and extracts all of the search criteria information, such as the names of people, and stores it in a database. The search system would then use the database to locate and extract all the transcription documents that meet the user's search criteria.