Validation of the CIDOC CRM Using both Extended Graphical and Category Theory Representations: Includes two New Zealand Case Studies

A Thesis presented in partial fulfilment of the requirements for the degree of

Master of Business Studies

in

Information Systems

at Massey University, Palmerston North, New Zealand

Jia Zhou

2004

Abstract

There is considerable interest in the use of the Internet to retrieve and integrate multimedia information from centres of cultural heritage such as museums and art galleries. The ultimate desire of most devotees of cultural matters is to have universal access, through a single portal, to detailed information from sites throughout the world. This level of interoperability is not an easy task both technically and culturally. To provide an avenue where some of the technical problems of accessing information from a huge range of unique database environments can be resolved, a semantic conceptual reference model (CRM) was proposed by The International Committee for Documentation of the International Council of Museums (ICOM-CIDOC). The model provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation. It is intended to provide a common and extensible semantic framework to which any cultural heritage information can be mapped. In this research two methods are proposed and developed to support the validation of the Conceptual Reference Model. The methodologies, one graphical and the other based on category theory, are used to replicate three published international validation activities and two new validations based on information supplied by two New Zealand heritage sites. This report also includes a literature review describing the main ideas and structures that form the basis of the CRM.

Acknowledgements

This has been an interesting but complex research activity. I would like to thank Judith Taylor of the Suter Galley in Nelson, and Susanne Geiser and Lynne Trafford of the Te Manawa Museum and Science Centre for their support and supply of records of cultural artefacts. I would also like to thank Bill Vernon and Abby Turbott of Vernon System for their assistance not only with this thesis, but research carried out at the honours level. I would like to thank Matthew Jackson for his support in applying Category Theory in this thesis.

I must not forget the support provided by the Information Systems Department and its teaching, technical and general staff.

I give my thanks to my friend Judy Engelbrecht for her support and friendship, my husband for his help on the format of this thesis and my family for their support and understanding during thesis writing.

Finally, I would like to express my thanks to my supervisor Barry Jackson for his support, encouragement and patience during the whole process of my research thesis.

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Background

The World Wide Web has transformed the way information objects are distributed and presented. Generally speaking centres of cultural heritage, primarily museums and galleries, have been quick to take advantage of the new technology, even to the extent that many of them manage their own web sites. Most of these sites tend to be simply showcases providing little more than links to a number of images, yet it is well known that what is displayed is only a very limited view of what is stored within collections. Information remains hidden from the public view, not because of a lack of enthusiasm on the part of the centres but because of the perceived difficulties in harvesting the information in a coherent and informative manner.

Some museums and art galleries have begun to investigate ways of exploiting the web as a global resource of cultural heritage information. Some have chosen to adopt an agreed metadata format (e.g. Dublin Core), while others have sought to create a 'universal' data model against which existing databases could be mapped. However, whatever approach is adopted, museums will have to establish solid and reliable systems that support the integration and distribution of rich and varied information contained in their collection systems.

1.2 Importance

Many observers believe that it is a mark of a civilised society to have access to objects of cultural interest and value. Cultural artefacts are clearly not located in one museum, or even one country or continent; they are dispersed in a variety of environments around the world. Gaining access and integrating associated information from such complex and dispersed environments requires electronic interoperability. The current situation regarding interoperability is one of uncertainty with a number of concepts and ideas under investigation by numerous research organisations throughout the world. Given the broad spectrum of approaches and ideas currently being researched, focussing on one contemporary approach could be seen as sensible and practical. It is with this pragmatic view in mind that the scope of thesis is confined to the object oriented semantic modelling approach pioneered by the CIDOC (International

Documentation Committee), a research group of the International Council of Museums (ICOM).

1.3 Issues

The major issues facing interoperability between centres of cultural heritage is the semantic and structural incompatibility of existing systems. Internationally based institutions have organised and structured their data in a number of different and sometimes unique ways. Even with the same collection management system (CMS), museums may have chosen to name and arrange entities and fields in quite different ways. Given the interests and values of individual archivists, it is quite likely that different levels of detail have been used to describe their collection. Even if the structures are compatible, terminology is often incompatible. The majority of solutions to this problem of incompatibility have been based on local transformation rules, or have adopted minimalist systems consisting of core data, and as a result have lost much of the richness contained in the original information.

For potential users of the CIDOC CRM the apparent complexity of the model and how this model might be mapped to existing data structures are important issues. These issues are addressed, in part, within this thesis.

1.4 Research problem

The research in this thesis centres on the object oriented Conceptual Reference Model developed by the document standards group of the ICOM/CIDOC. The aim of the CRM is to provide ways for museums to render their information resources to one another without losing detail or precision.

The problem faced by this researcher and many archivists and information systems professional wishing to understand and work in this field is the complexity and utility of the CRM framework.

It is intended in this research to address this problem in four ways:

- Undertake a literature review bounded by the CRM perspective
- Develop tools to enable researchers map archival data to the CRM
- Use the new tools to replicate the CRM validation exercises published in the international research literature

- Apply the new tools to two New Zealand centres of cultural heritage.

1.5 Aims and objectives

The aims of the research are:

1. To gain an understanding of the CIDOC Conceptual Reference Model (CRM) used to represent the semantic content of cultural data held within museums and art galleries.

   Note: The scope of this objective is primarily limited to published research undertaken by The International Committee for Documentation of the International Council of Museums (ICOM-CIDOC) and the Institute of Computer Science of the Foundation for Research and Technology – Hellas (ICS-FORTH) under the direction of Martin Doerr.

2. To develop new tools to assist in the validation of the CRM against real-world heritage collection systems.

   Note: An extension to the graphical tool used by Martin Doerr is proposed as well as a new mathematical tool based on category theory notation.

3. To investigate, using the above mentioned graphical and mathematical notation, several of the seminal publications used to validate the CRM on an international level.

   Note: The international cultural heritage systems being; Encoded Archival Description (EAD), Dublin Core (DC) and Art Museum Image Consortium (AMICO).

4. To apply the same graphical and mathematical notation to validate the CIDOC CRM within the New Zealand context.

   Note: The two New Zealand centres of cultural heritage being; The Suter Gallery in Nelson and Te Manawa, a Museum and Science Centre in Palmerston North.

1.6 Research process

In general terms the process adopted in this research follows that proposed by Bourner (1996, p7). There are four steps in Bournier's approach:

1 Research the field of study

2 Develop a model or framework

3 *Test the model*

4 *Undertake evaluation and reflection*

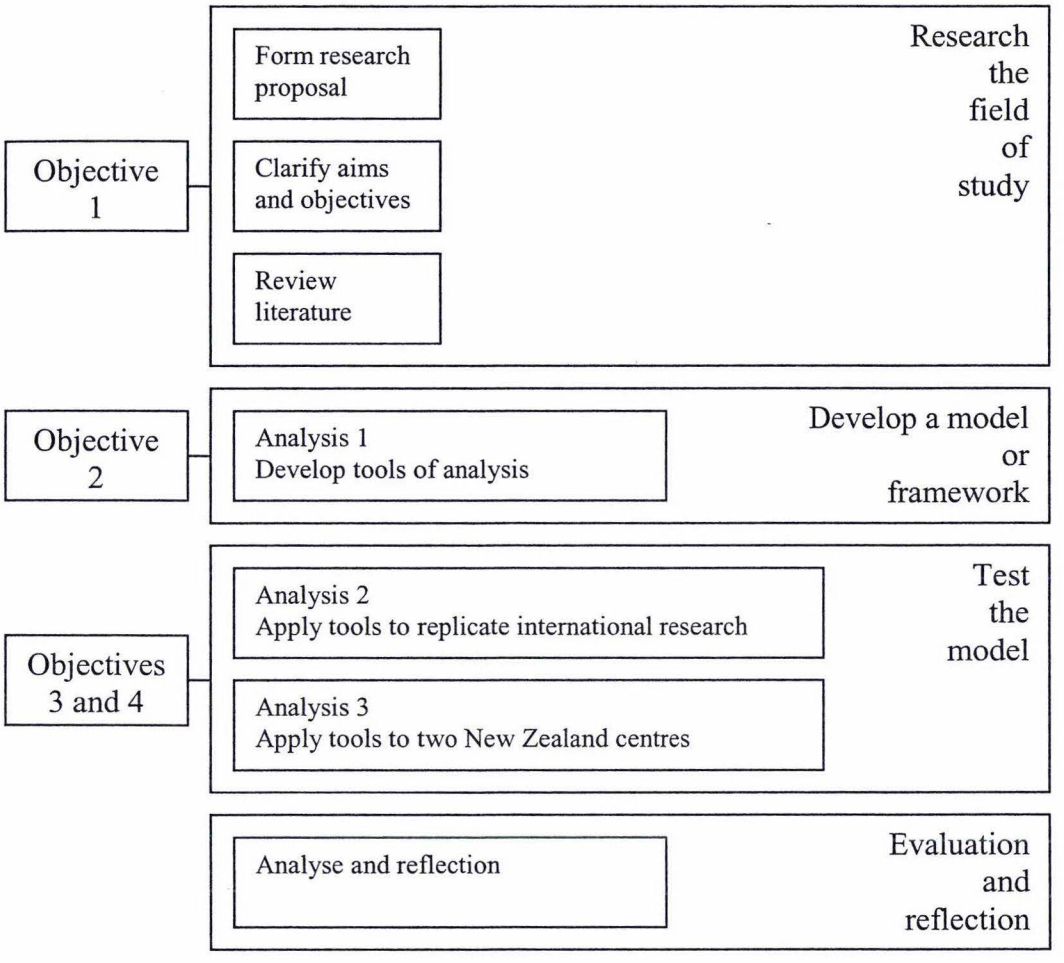


Figure 1 Research Process

1.7 Report structure

The report has the following structure, which is aligned with the Aims and Objectives set out in Section 1.5 and the Research Process in Section 1.6.

Chapter 1: Introduction

The rationale for undertaking this research is outlined in this chapter, together with a description of the research problem and lists the main aims of the research.

Chapter 2: Research design: Methodology

The 'four step research approach' proposed by Bournier provides the guiding framework for the research design. In fact, the research processes are aligned not only to the research process but also to the structure of this report.

Bournier's steps		Research Report	
		Chapter Numbers	Chapter topic
1	Research the field of study	3	Literature Review
2	Develop a model or framework	4	Validation Tools
3	Test the model	5	International validation
		6	New Zealand validation
4	Undertake evaluation and Reflection	7	Conclusions

Chapter 3: Literature Review (Introduction)

The chapter is concerned with discussing the origin of the CIDOC Conceptual Reference Model, in particular why such a model is needed to facilitate the sharing of cultural information.

Chapter 4: Analysis 1: Validation tools -- Graphical and Mathematical Notation

The need to view the validation exercises in a consistent manner was the motivation to develop both graphical and mathematical representations.

Note: There is a possibility that applying category theory to the CRM domain could lead to an effective parsing algorithm to link cultural databases to the CRM. Such an opportunity is outside the scope of this research.

Chapter 5: Analysis 2: International Validation of the CRM

This chapter brings to bear the graphical and mathematical notation, developed in Chapter 4, to explore and replicate three representative investigations undertaken by researchers to validate the CRM.

Chapter 6: Analysis 3: New Zealand Validation of the CRM

Essentially this chapter applies the techniques used in chapter 5 to two New Zealand centres of cultural heritage.

Chapter 7: Discussions, Conclusions and Recommendations

The final chapter provides a summary of the findings of this research and discusses their relevance to the main aims stated in Chapter 1.

Bibliography

All the references emerged in the report are listed in this section.

Appendices

The appendices contain information from the CIDOC CRM. The data are only a small sample of what is provided in the original CRM documentation. It provides reference documentation to support the validation processes described in this thesis. Care needs to be exercised as the CIDOC CRM has been modified several times in the recent past and different versions are referenced in different sections of this report.

2 Research Design

2.1 Introduction

There is an appreciation within the international community that museums and art galleries have stored in their collection systems vast quantities of information relating to cultural data. Regretfully, there appears to be no real and effective mechanism for that information to be retrieved, integrated and displayed over the Internet of viewed via other electronic means. Several attempts at seeking a universal data model have been made (Doerr, 2001b) that could be used to 'capture' the data, but most, if not all, have been abandoned. Martin Doerr and his associates at the Foundation for Research and Technology – Hellas (FORTH) proposed an innovative and unique approach to this complex problem of interoperability. The main outcome of their research was a semantic model that could be used to reference the cultural information captured and stored within centres of cultural heritage. In effect, the aims of this research are to analyse and build on the body of knowledge produced by the FORTH group.

2.2 Methodology

As previously mentioned the journey travelled by the FORTH researchers is one focus of this research. In addition, the researcher intends to develop new ways to view the validation of the CRM undertaken by international groups and apply them within the New Zealand context.

When developing the research methodology for this research activity, the researcher was keen to adopt the four-step approach to research proposed by Bourner (1996, p.7).

1 *Research the field of study*
2 *Develop a model or framework*
3 *Test the model*
4 *Evaluate and reflect.*

2.2.1 Research the field of study

Researching the field of study in this research project is divided into two main stages. However, it is true to say that reference to learned documents are to be found throughout the report.

In stage 1 (Chapter 3: Literature Review) a review in undertaken drawing upon the published work of the FORTH researchers and others, relating to the foundation of the CIDOC CRM. It draws upon earlier research that examined some of the core concepts that underpin the development of the CIDOC CRM.

Stage 2 (Chapter 5: Analysis 2) This is a combination of a literature review and a validation exercise. Three published papers, which seek to demonstrate the ability of the CRM to capture information stored in three international and established metadata structures, are reviewed and the processes replicated using mapping documentation tools.

2.2.2 Develop a model or framework (In this case a graphical tool and mathematical documentation mapping tool)

In this step (Chapter 4: Analysis 1) two ways of describing and documenting the CRM validation process are proposed.

- A graphical representation, which has its origins in the FORTH environment. The notation used by Martin Doerr (2000) has been modified in a number of ways to allow for a richer range of situations to be expressed and displayed. The graphical notation shows how the various CRM entities are linked together by unique properties. A brief literature review is provided illustrating the importance of using graphics when 'use' and 'structure' are being represented.

- The form and structure of the CRM lends itself to representation using category theory, as a consequence a new mathematical notation is proposed.

2.2.3 Test the model (in this case the CRM and the graphical and mathematical filter)

The two documentation-mapping tools are tested in this, the third step of the Bournier's approach (Chapter 5: Analysis 2 and Chapter 6: Analysis 3).

Once the CRM model was near completion, Martin Doerr sought the support of the international research community to validate the CRM against 'real-world' records of cultural artefacts (Doerr, 2003). In this research three published international validation processes (Chapter 5) and two New Zealand sites of cultural heritage (Chapter 6) are examined

2.2.4 Evaluate and reflect.

The final step in the Bournier research process, evaluation and reflection, is to be found in Chapter 7.

3 Literature Review

3.1 Introduction

The literature review is distributed across several chapters in this report. The purpose being to place the published research more appropriately within the report, allowing detailed review and analysis to take place using the special graphical and mathematical tools developed during the research process. Justification for this approach is that understanding the CRM semantic model and its associated structures are seen as key objectives of this research endeavour.

- Chapter 3: Literature Review (Initial)

  In this, the current chapter, the need for new approaches to dealing with the complex problem of the interoperability of centres of cultural heritage is examined. In particular the CIDOC Conceptual Reference Model.

- Chapter 4: Analysis 1: Diagrammatical and Notational Representation.

  A brief review of the literature supporting the use of graphics and category theory is undertaken.

- Chapter 5: Analysis 2: International Validation of the CRM

  Three publications describing the validation process for data held in three established cultural record formats are reviewed. Each journal publication is examined using an integrated combination of a graphical representation and mathematical notation. The original format of the original publications is retained to ensure comparability and consistency. This approach has a dual-purpose as it not only develops a better understanding of the validation process but it also provides confirmation of the utility of the two notational tools.

3.2 Needs of Interoperability of Cultural Heritage Information

3.2.1 Introduction

There has been a growing interest by museums and art galleries to display images of cultural artefacts on the Internet. This is not limited to large international organisations, more modest institutions such as those found in New Zealand already display some or part of their collections online. New Zealand Museums (http://www.nzmuseums.co.nz/) provides links to almost all museums in New Zealand;

it also provides a search function, which people can use to locate specific collection types. Te Papa, The National Museum of New Zealand in Wellington (http://www.tepapa.govt.nz/) and Te Manawa in Palmerston North (http://www.temanawa.co.nz), have placed images of some of their collections online.

The public and professional researches are looking for much more and would like to see a single portal that would allow a single search query to poll and access collections from numerous centres of cultural heritage worldwide. Coupled with this universal access, the public would like to see much more of the detailed information that is known to exist describing the various objects. This might include text, graphs, animation and video.

3.2.2 Challenge of information integration

With the development of the World Wide Web, information is being gathered and distributed worldwide. Many museums and art galleries are seeking to adapt their systems to take advantage of the new technologies, and the use of on-line technology is providing them with a powerful tool to meet this challenge. However, Doerr and Crofts (1999) complain that many museum websites are simply offering a quick and limited look at the source information that is available and ignoring the enormous amount of information collected by museums, instead of putting effort to integrate their resource with other institutions. Doerr and Crofts (1999) suggest that in order to reach the vision of integrating and distributing detailed resources, museums need to establish solid and reliable means, and integrated structures.

Access to museum information has the potential to attract a wide international audience, such as the general public, researchers and educational institutions. Doerr and Crofts (1999) point out that integrating information from many different sources is essential and has the potential to enhance the value and richness of the information. It would not be unrealistic to expect that the integration of contextual information available across different institutes could actually promote interest in cultural information and generate a greater awareness of our's and other societies.

Many researchers and groups share this vision and are examining ways to integrate and distribute global resources for cultural heritage information. For example, the Digital Library Research Group at Cornell University and the Corporation for

National Research Initiatives (CNRI) have been engaged in research that focused on the design and development of infrastructure for an open architecture, confederated digital library (Payette and Blanchi, 1999). Gruber's (1993) research on exploring the use of formal ontologies as a way of specifying content based information for sharing and reuse of knowledge is another initiative. To some extent the CIDOCC CRM researchers built upon the work by Gruber.

Based on research conducted by Doerr and Crofts (1999), the challenge of information integration and distribution is to overcome the incompatibility of the semantic information and data structure of the many existing museum systems. This challenge is primarily caused by the differences and variety of museum databases, which organise and present their data structures in different ways. The differences in the data structure may also lead to incompatibility of the naming and arrangement of the entities and fields. Doerr and Crofts (1999) make the further comment that even if the database structures were compatible, there would still be a problem of incompatibility of the terminology and language. Doerr and Crofts (1999) conclude that researchers need to address the problem of these incompatibilities, regretfully; many of these attempts are based on hermetic and special transformation rules, or just simply concentrating on a limited subset of 'core' data.

3.2.3 Need for semantic interoperability

As Doerr (2001b) points out, the development of the World Wide Web increases the possibility of data transference, comparative studies and data migration between heterogeneous sources of cultural contents. Doerr (2001b) claims that there are dozens of "standard" and hundreds of proprietary metadata formats existing in the cultural area. One such system, the Dublin Core is considered by some to be too limited and unable to meet the advanced requirements of an increasingly demanding and informed target audience. Information invariably becomes diluted during the transition from the original source to the Dublin Core. In Doerr's (2001b) research, he claims that many of the data and metadata structures place more emphasis on optimising coding, storage cost and data structure for specific application than on the value of information. They tend to be designed for data capturing instead of showing the meaning of the contents, and this results in a flat data structure and some of the meaning is unavoidably lost or hidden in these structures.

Doerr (2001b) argues that simplifying the data structure without losing the meaning is essential if interoperability is to be achieved. Doerr discusses two scenarios – neither of which provides a complete solution.

1. Reduce complexity – is it possible?

   Doerr points out the 'complexity' paradox - in order to guide users to the enormous information and data, the data structure needs to be complex to accommodate the use of formal queries. However, as Doerr points out, the more complex the system becomes, the more difficult it is for the user to handle, more expensive for the user to afford, more time is required for the user to learn to use the system, and more time is required to respond to the request and deliver the content. According to Doerr, most museum applications are running in unique and dedicated environments, such as a library, art gallery or historical archive, and because of the complexity of these applications, it is unlikely, even impossible for these applications to be integrated with each other. Doerr claims that in order to create efficient applications, simplification of the data structure becomes necessary or a new solution needs to be developed, such as a semantic model.

   Note: Doerr claims that one of the key reasons why the Dublin Core is so popular is its flat structure. However, Doerr claims from his experience the Dublin Core is not simple to use!

2. "Finding aids":

   Another reason to simply data structure is that it will facilitate the use of the "finding aid". Simplification of the target data source is required as even modest differences in the data structure from one environment to another can significantly affect the effectiveness of a finding aid. Doerr claims one of the justifications for using a "flat" metadata schema such as the Dublin Core is to present to the finding aid a consistent and generic structure. Although the Dublin Core simplifies the data structure by grouping data, it is at the expense of losing relationship information between elements of one group to elements of another. Doerr suggests, while it is necessary to improve the data structure, it is still a requirement of the finding aid to "recover" the hidden meanings.

The situation described is shown schematically in Figure 2. As shown in the figure, the three circles represent three different data structures in three different environments. User 1 attempts to receive the information from the three different sources, he is very confused and frustrated about the different data structures making retrieval complicated. User 2 gets the information from the three places, where data has been mapped to the Dublin Core. His situation is better than User 1, however, the data provided is limited by what the Dublin Core structure can provide. User 3 is in the most desirable situation, as he receives all the information as it is mapped to a generic semantic structure without modifying the content or structure of the original data. The use of a semantic approved not only ignores the complexity of the data structure but also maintains the integrity of the information.

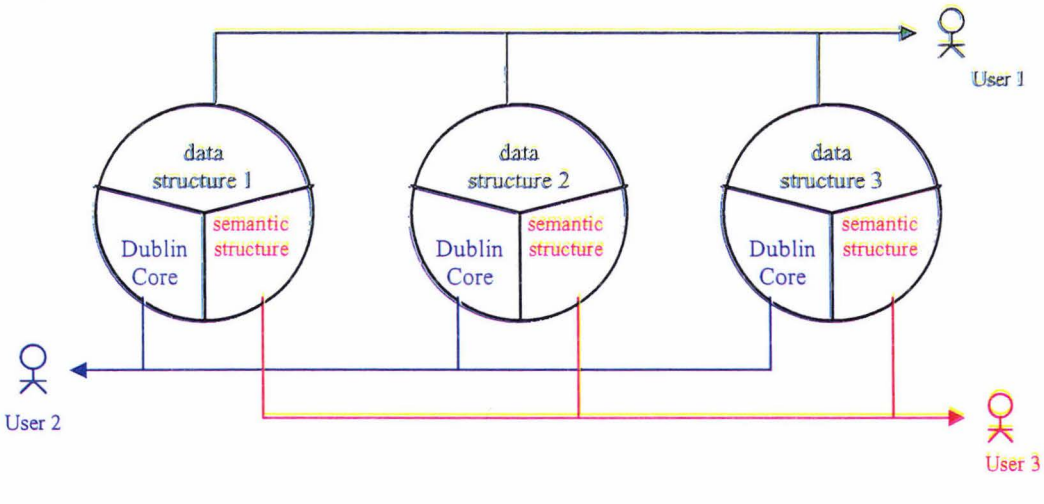


Figure 2 Comparison of accessing various database structures

3.3 *Foundation of CRM: the Use of Ontology*

3.3.1 Introduction

The CRM has its roots in the field of ontology. In the following section, some of the foundation concepts and ideas are explored.

3.3.2 Ontology

Gruber (1993, p1) provides a definition of an ontology, "*an explicit specification of a conceptualization.*" According to him, "*conceptualization is a conceptual, abstract view of the world of what people would like to present for a particular purpose.*" He goes on to say that, "*...the body of the ontology is based on a conceptualization.*" Gruber (1993) states that an ontology can be used to describe, by defining a set of

associated names of entities, such as classes (entities), relations (properties), functions or other objects, textual information that is readable by human beings.

Gruber (1993) proposed a set of good design criteria for constructing an ontology whose aim is to share and integrate knowledge based on a shared conceptualization, these criteria are:

1. Clarity

   An ontology should express the meaning of all defined terms in a clear and effective way. All definitions must be formalized and independent of any social or computational text. Gruber also points out that all the definitions should be documented in a natural language.

2. Coherence

   An ontology should be consistent with the definition and also coherent with the natural language used to define it.

3. Extendibility

   An ontology should be designed to be flexible enough to anticipate any future changes. That is to say it gives the opportunity for users to define new terms without having to revise existing definitions.

4. Minimal encoding bias

   The conceptualisation should be specified at the knowledge level without depending on a particular symbol-level encoding. According to Gruber, an encoding bias might occur when a term is chosen purely as a matter of convenience of notation or implementation.

5. Minimal ontological commitment

   An ontology should achieve the meaning contained within the knowledge base by using the minimal ontological commitment. In other words, minimal ontological commitment can be achieved by including only those terms that are essential to represent the intended meaning. This implies that there is a consistent use of a vocabulary. The following diagram, by Guarino (2003), illustrates this point particularly well.

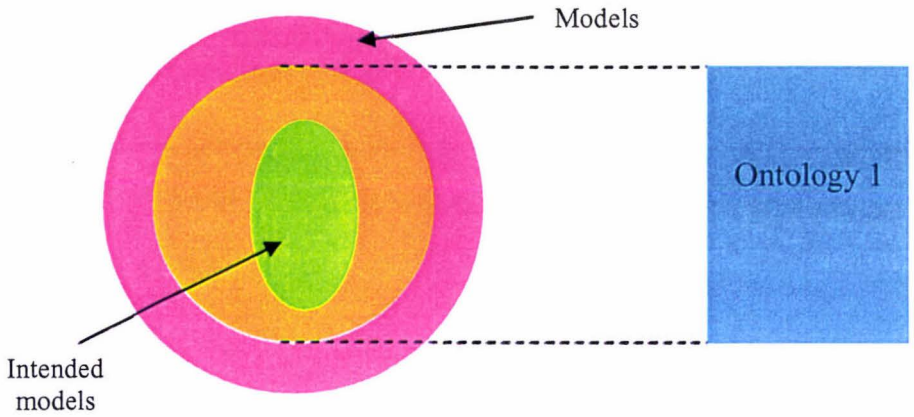


Figure 3 Poor Minimal Ontological Commitment (Guarino, 2002)

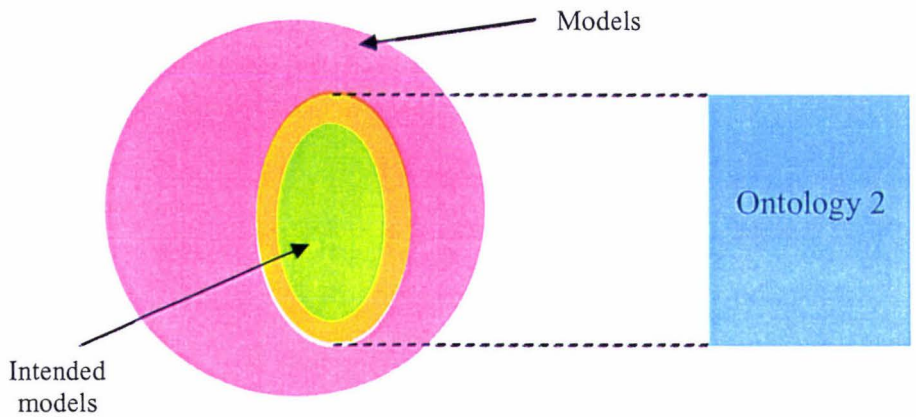


Figure 4 Good Minimal Ontological Commitment (Guarino, 2003)

In Figures 3 and 4, the outside pink circle represent all the available models that include the 'conceptual specification' of the intended model; the middle yellow circle represent the ontology considered to be the best way for embodying the meaning of the intended model; the inside green one represent the intended model. Comparing the two ontologies in Figure 3 and Figure 4, it is quite obvious that the first ontology produces a surplus of terms when representing the intended model. The structure of the second ontology looks more precise and is able to map the intended model by using a minimum of terms considered essential to represent the meaning of the knowledge.

6. Tradeoffs

The previous five criteria may sometimes contradict each other under certain circumstances. The most likely and obvious contradiction is between clarity and ontological commitment. To improve clarity, terms are more tightly defined, however, this may; as a result, require a loosening of the minimal ontological commitment. The clarity criterion refers to how well terms are defined, whereas ontological commitment refers to the conceptualization being described. The analyst, when establishing the ontology will need to make a decision on which option they would like to take in order to achieve the best compromise.

3.3.3 Uses of Ontology in Information Systems

Ontology has been widely used in information systems. Guarino (1998) points out that research on the use of ontologies has continued to increase in the computer science domain. Guarino (1998) defines ontology as:

> "*In the simplest case, an ontology describes a hierarchy of concepts related by subsumption relationships; in more sophisticated cases, suitable axioms are added in order to express other relationships between concepts and to constrain their intended interpretation.*"

Guarino (1998) claims that ontology is becoming one of the most popular methods used in the development of database components. Guarino (1998) shows that during the development of a database, the final conceptual model can be displayed as a computer processable ontology, which can be mapped to the principal target platform. According to her, these aspects have been extensively studied for the mapping of "knowledge specification" to schemes for many different types of databases, such as relational, object-oriented, deductive, active, etc. Guarino (1998) claims that utilization of a highly interdisciplinary approach is the main peculiarity of the methodological side of the ontology, and this peculiarity is seen to be one of the most important features of CRM integration for cultural heritage data and information.

Guarino (1998) also describes the use of ontology during the development phase as "ontology-driven IS". She then points out several benefits of using ontology at development time: it enables the developer to practice a "higher" level of reuse; it

enables the developer to share the domain knowledge using a common vocabulary across different software platforms. In fact, CRM can be regarded as a typical example of "ontology driven" information systems. Gruber (1993) has a similar description of ontology; he suggests that a common ontology defines the vocabulary and the sharing of that vocabulary on different agents.

Gruber (1993) also mentions the possible use of ontology in the bibliographic domain. According to him, in a bibliographic ontology, data fields such as name and date are in the format of the bibliographic domain. Format for referencing author's name or date must be specified according to the concept of certain kind of format in a specific historical time. However, problems may occur in a bibliographic ontology. One of the examples given by Gruber (1993) is the use of dates in the bibliographic domain. The date domain has its own standards for the precision in which time is measured. For example; the Chinese calendar uses a different system from the Western calendar so this must take this into consideration when referencing time. However, questions might arise as to whether the specification of a standard measurement or identification scheme contradicts the design criteria of "minimal encoding bias" and "extendibility". Gruber (1993) explains that it does not generate encoding bias or limit extendibility. He points out that the notion of time is an independent unit, and the unit is introduced by the mapping from the time point to the surface encoding. This enables an agent working on the Chinese calendar to read the date using the Western calendar, and the agent subsequently encodes it in the appropriate format.

3.3.4 Features of CRM Ontology

Described by Doerr (2001b), CRM is a high-level ontology that allows the joining and sharing of cultural heritage data by accessing library and archival information. When describing CRM, Doerr defines CRM ontology in terms of computer science rather than philosophy. He expands on this to claim that the CRM ontology is an approximation of a conceptualisation of a domain supported by a formal language and a vocabulary. The link with Gruber's work is clear.

In this section of the report, the major features of the CRM ontology are discussed. These discussions are based primarily on Martin Doerr's research.

3.3.4.1 Integration of Context-Free propositions

According to Doerr (2001b), the CRM Group's vision was to create a global semantic network model. The purpose of the model was to facilitate the joining together of cultural knowledge from all archival sources. No single museum is able to obtain and store all the records about a given subject and its associated artefacts. The desired museum 'data' structure should not attempt to create a single view of an object, but to maintain information links to all the sources of relevant information.

Doerr (2001b) produced a set of properties that he considered essential for good ontology design. Some of these terms appear to have their origins in work undertaken by Gruber (1993).

1. Context-free interpretation

   Each of the statement in the CRM ontology model should be interpretable without knowing any contextual data. Doerr (2001b) states that the global identifiers are "fix point", and that information around the global identifiers can be interpreted directly without the need for any other process. Based on Doerr's (2001b) description, context-free interpretation is achieved by putting the global identification of individuals on one side of the relationship, and then put the appropriate design on the other side. Such as $(has\ a)$: $birth\_date$ does not make sense without another entity such as "creator".

2. Alternative views

   Doerr (2001b) claims the CRM model should be able to capture multiple alternative statements or viewpoints about any so called 'fact'. In a way this demonstrates the complexity of cultural information – and for many experts is the very essence of human society. The compilation of alternative statements in well-defined points of the semantic model is a great help for users and any subsequent reasoning that might take place.

   Example 1: A frequently referred to example about compilation of alternative statements given by Doerr (2001b) is the Union List of Artist Names. This list comprises 'life' data of more than 100,000 artists. The data is obtained from numerous sources and the opinions of experts are often found – sometimes there are opinions on opinions. As one would expect from a diverse set of people, some of these opinions may be contradictory.

Example 2: When collecting the relevant information about the Korean War, one may find that North and South Korea have different understanding of events. It will be very helpful for the user to access and view the information in a structured and meaningful manner. Allowing users to gain a much richer and perhaps more balanced view of events.

3. Appropriate granularity

Some artefacts are extremely well documented while others are short on detail. The more detail is encoded about an object, the greater the degree of granularity.

Example: In order to have a more explicit view of information, it is necessary to "dig out" the hidden concepts of the model. Doerr (2001b) gives an example about documenting 'related' information about the birth of an artist; it is not sufficient to mention only the usual properties, such as "birth_date" and "birth_place", in his view, there are other hidden pieces of information that could be interesting and informative. The CRM structure explicitly permits hidden concepts to be expressed. This is clearly not the case with such schema as the Dublin Core, which has a low level of granularity.

4. Principle of "minimal ontological commitment"

Doerr (2001b) supports Gruber (1993) and Guarino's (2003) opinion that an ontology should endeavour to support meaning using the minimum number of entities and properties. It is important to appreciate that the CRM is designed to express information from all forms of cultural objects whether they be man-made or otherwise. If we were to limit the CRM to man-made objects then the model would require fewer entities and properties to express the information. Similarly, as new forms of information come on stream, for example through the use of more complex forms of media, then the CRM might need to be enlarged. This is mentioned in the next section.

3.3.4.2 Monotonicity (Extendibility)

Gruber (1993) claims that terms should be designed to be flexible enough to anticipate any future changes, and provide users with the opportunity to define new

terms without having to revise existing definitions. Gruber (1993) expresses this feature, "*So one can extend and specialize the ontology monotonically*".

1. Classification and specialization:

   Doerr (2001b) states that once an object is classified according to the current state, it may also be classified by a subclass of itself, this is so-called multiple instantiation.

   Example 1: CRM has a hierarchy structure, and the hierarch level may need to be changed or amended if the state of a certain object is changed. Based on the "classification and specialization" rule, once the object has been classified, no matter what changes are made to the object, its subclass cannot change the previous definition.

   Example 2: Doerr (2001b) gives an example of preservation of classification. Some large Minoan terracotta vessels have been found in Crete. Due to their similarity with modern bathtubs, they were initially regarded as bathtubs until, much later, bones were found inside one of them. They were then recognized as sarcophaguses. Because the object was initially defined as a container rather than a bathtub, the additional knowledge found out later would not have invalidated the previous one.

   Doerr (2001b) summarizes this feature, "*an ontology which preserves classification of its instances under addition of non-contradictory knowledge is also monotonic under extension of its class system.*"

2. Attribution

   In CRM, one entity can be directed to various attributes; in Doerr's words, this is a "richer" system. Doerr (2001b) states such property paths are potentially infinite, however, in a "poorer" system, the number of attributes is restricted perhaps limited to just one.

   Example 1: When displaying the information about a physical object, if the "poorer" system is adopted, the physical object will be directly linked to its condition state using the terms say, "good" or "bad". However, in a "richer" system the intermediate entities will not be omitted and perhaps several condition states might be expressed; rough texture, heavy, green, .. etc.

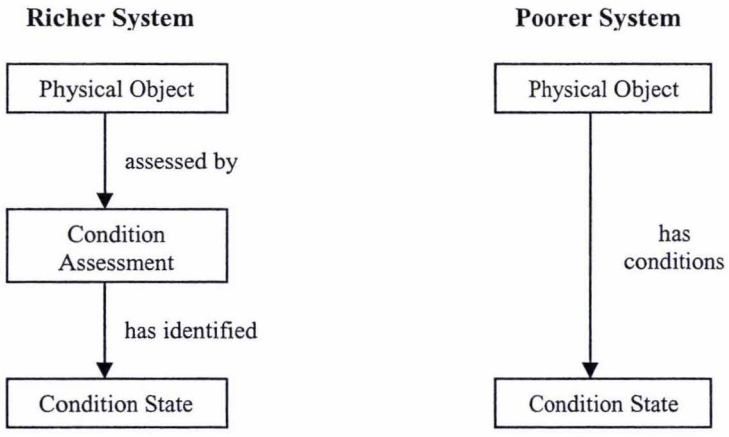


Figure 5 Richer System and Poorer System – based on Doerr and Crofts (1999)

Example 2: A house has been rated as first class, a "poorer" system may not refer to the reasons and the specific assessments of the outcome, just simply register a rate of "first class". However, if it is a "richer" system, it may refer to the detailed condition of the house as an assessment of the outcome of a number of measurements carried out by specific people over a period of time.

3. Alternative models

According to Doerr (2001b), the ontology monotonicity can be achieved by using different modelling alternatives. This point is illustrated in the following examples:

Example 1 Avoiding unconfirmed states: Doerr points out that the state of a phenomenon can change as time passes. Doerr suggests that if the information is not complete the transaction state cannot be recorded. According to him, status is easier to observe than state, such as the validation of an object at a certain points in time. Based on the above consideration, Doerr (2001b) points out that in CRM, ownership changes is preferred to recording ownership states.

Example 2 View-neutrality: Doerr (2001b) explains that for a museum, registration involving the transfer of an object (and/or its record) from one museum to another is treated as a deaccession event for one museum and accession event for the other. Doerr points out the classification of "deaccession" and "accession" are regarded as non-monotonic. He suggests that it would be better to replace these two notions by using one of the symmetric terms in CRM, such as Acquisition or Change of Custody. However, such a change of focus may not meet with the support of museum staff!

3.4 The CIDOC Conceptual Reference Model

3.4.1 Who is involved?

The first version of the CIDOC Conceptual Reference Model (CRM) was created by CIDOC's Documentation Standards Working Group (DSWG) in 1999. The focus of CIDOC (International Document Committee), arguably one the most influential committees in the museum industry, is to ensure that the documentation interests of museums and similar organisations are appropriately managed and implemented. The Committee is one of 25 component international committees of the International Council of Museums (ICOM), established during the 1950 ICOM General Conference in London. CIDOC and the CIDOC Documentation Standards Working Group (DSWG) had previously been engaged to represent cultural data in the creation of a general data model, with a special interest on information interchange. The general data model approach was abandoned in 1996 in preference to a conceptual semantic model. The data model had grown out of all proportions in its attempt to meet the needs of the various supporting organisations. It was seen as unworkable. The idea of creating a Conceptual Reference Model was initiated by DSWG who adopted an object-oriented approach in 1996; the aim of the model was to focus on getting benefit from its power and extensibility for dealing with the necessary diversity and complexity of museum data structures. The CIDOC CRM was accepted as a Committee Draft by the International Standard Organisation (ISO) in December 2002. According to Doerr (2003), the CRM is in a very stable form. And it is now registered as ISO/CD 21127 and is expected to become an ISO standard in 2004 (Doerr, 2003). However, researchers and specialists are continued to seek to improve the theoretical understanding of the model, and currently several applications and comparison of the model are still underway. In fact, Doerr (2003) specifically requested international organisations attempt to validate the CRM against local and international collections. This process is still underway and it has become one of the key aims of this research.

3.4.2 Introduction of CIDOC CRM

CIDOC CRM is a high level ontology designed to provide definitions and formal structure that can be used to describe the implicit and explicit concepts relevant in the area of cultural heritage (Doerr, 2001b). The CRM provides a semantic structure that allows the mapping of different sources of cultural heritage information with the express purpose of promoting a shared understanding of cultural heritage information.

By providing the so-called "semantic glue" suggested by Doerr, the CRM is able to provide access between different sources of cultural heritage information. As Doerr (2001b) concludes, CRM is an ontology formulated in the form of an object-oriented semantic model that aims to solve the problem of semantic interoperability, enabling various kinds of related museum data to be accessed while maintaining the richness and complexity of the original sources.

3.4.2.1 Overview of CRM Structure

The CRM structure is the subject of this section. Part of CRM version 1.1 is shown schematically in Figure 6 and version 3.3.1 in Figure 7. To facilitate understanding of the CRM, a set of CRM terminologies is described, some of the basic entities are explained and examples of CRM models are presented.

1. Basic entities:

   It is a challenge to present the overall structure of CRM in a succinct and meaningful way. It is difficult knowing where to start, as there are so many different interrelating concepts and ideas. Doerr and Crofts (1999) suggest the use of a "top down" approach to examining the model. There are some inevitable drawbacks in this approach, as it may be difficult to grasp the practical application of the model as one is starting from such a high-level of abstraction. Figure 6 presents the main branches of the class hierarchy used by Doerr and Crofts (1999).

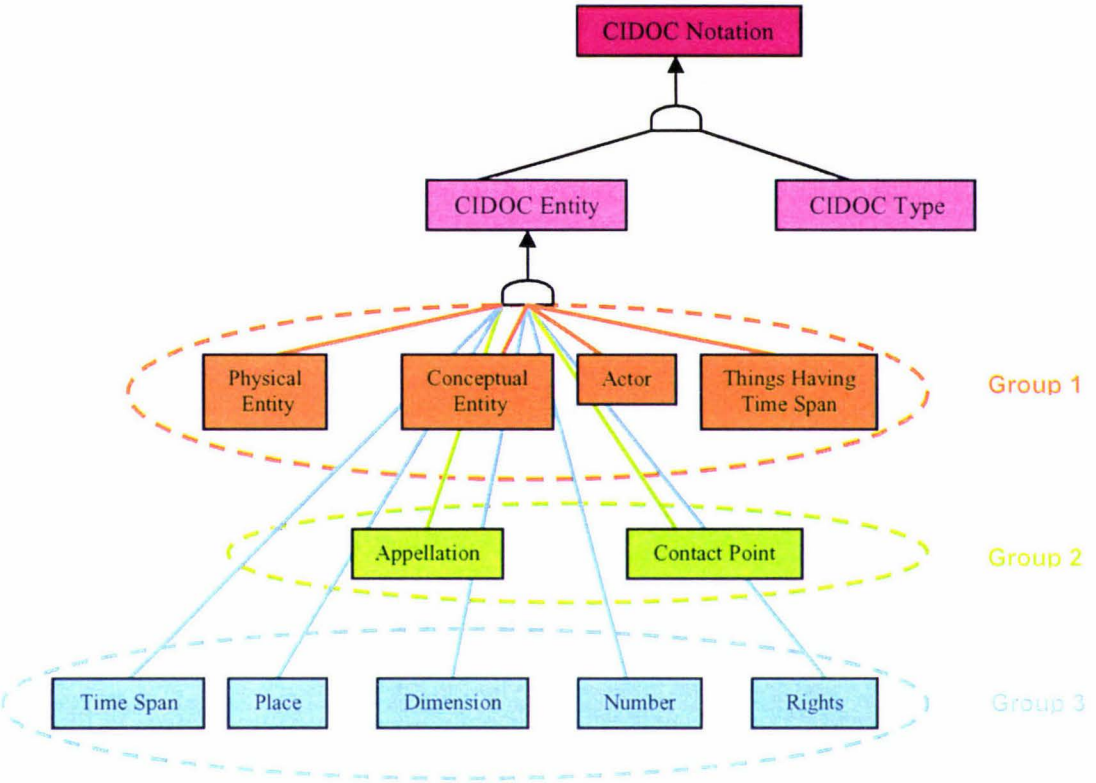


Figure 6 CRM Main Entities based on CRM Version -- based on Doerr and Crofts (1999)

In Figure 6, different colours represent different hierarchical positions of the entities. The symbol "  "is used to represent the hierarchical relationship between the entities

Detailed descriptions of the entities in Figure 5 follow; the ones presented here are based on CRM version 1. The CRM model has undergone significant changes over the years, however, version 1 illustrates the fundamental concepts and it is relatively simple to understand.

- **CIDOC Notation** is the highest level in the CRM, and it works as the notation container for all the other classes. According to Doerr and Crofts (1999), it has no significance beyond this, and can be ignored for most intents and purposes.
- **CIDOC Type** is an additional class. In the above diagram, "Type" describes the CIDOC Notation Entity, however, "Type" can be used to describe other classes in the same way. The CRM structure provides a mechanism for enhancing the level of detail. In other words, the CRM supports increased granularly and monotonicity.
- **CIDOC Entity** is the parent class for all the major classes in the CRM. Doerr and Crofts (1999) mention that the above diagram shows how, for presentation purposes, the subclasses are separated into three groups, the first group consists of four basic concepts that are fundamental to the model and focus

entirely on cultural heritage documentation. It covers the classes of Physical Entity, Conceptual Entity, Actor and 'Things having Time Span'. The second group consists of some subordinate concepts used in the CRM, this group covers the classes of Appellation and Contact Point; and the third group is a set of primitive classes which is used to describe attributes types in the model, typically in providing description of the basic entities.

- **Physical Entity** contains all the physical features and physical objects residing or documented in museum collections. This covers objects / features such as mountains, rivers and seas.
- **Conceptual Entity** is used for all intellectual or conceptual objects. Such as books (linguistic document as opposed to a physical thing), paintings (visual object that stimulates the mind) and agreements. The distinction between a conceptual and physical entity is like the difference between the concept of an agreement and its documentation, such as the Treaty of Waitangi and its relevant documentation/agreement of the Treaty. In this case a physical copy of the documentation would be kept on file. Doerr and Crofts (1999) extend the CRM classes to include other conceptual objects, such as Designs and Procedures, Linguistic Object and Visual Items.
- **Actor** includes the class of all agents, such as persons, groups and institutions. These are entities capable of action, and also those who are potentially responsible for an event taking place, in other words - capable of causing a change of state.
- **Things having Time Span** is the class concerned with periods, events, and states. In fact, all the forms which are inconstant in time.
- **Appellation** is the class consisting of all names, codes or words, either meaningless or meaningful. Doerr and Crofts (1999) state that appellation is used to identify an object by using a conventional or traditional format, or by agreement.
- **Contact Point** consists of all the contact information for agents or objects. The information covers addresses, telephone numbers, email, post office boxes, etc.
- **Time Span** is a combination of a set of dates or duration, which can be used to indicate a period of valid time, associated with an event or any other phenomena.

- **Place** is used to describe the areas in space, particular on the surface of earth. According to Doerr and Crofts (1999), places are usually identified by reference to large "immobile" objects, such as rivers, mountains, buildings, etc.
- **Dimension** class contains numerical values used to describe measurement taken on the objects. Doerr and Crofts (1999) consider currency, length, diameter, weight, weight, density, luminescence, and percentage of tin content as numerical values.
- **Number** consists of a set of mathematical numbers, considered as a data type.
- **Rights** contain all the legal rights for the objects, such as property rights, etc.

Figure 7 shows all the classes and their hierarchy relationship with each other. This figure is based one of the recent versions of CRM Version 3.3.2. (Crofts, et al, 2002).

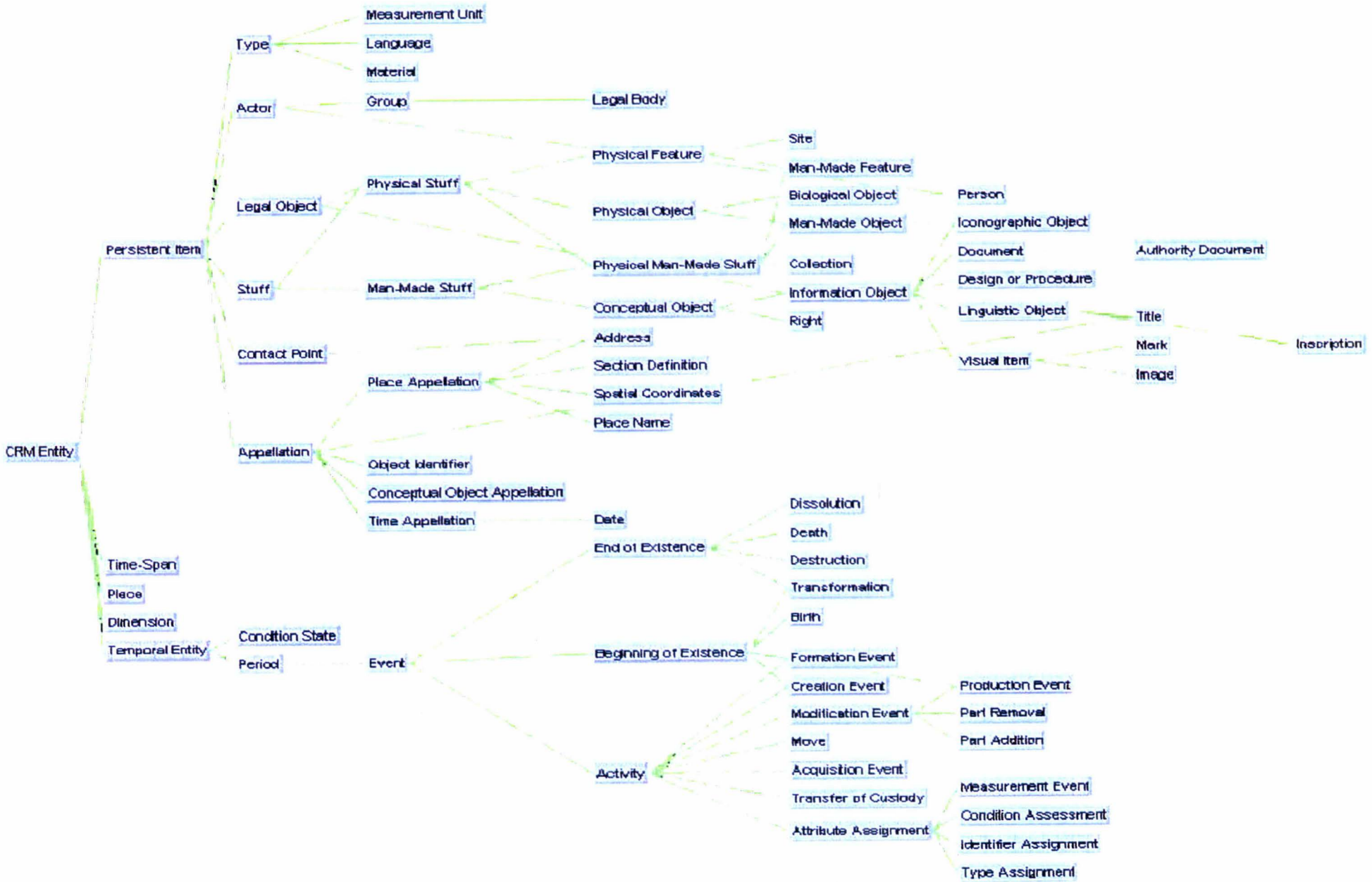


Figure 7 CRM Struture (Version 3.3.2)

2. Description of the CRM terminologies

Crofts, et al, (2001, pages iv-vi) provided the following definitions of some of the basic CRM terminologies (CRM Version 3.0):

Entity: *"anything that may be called 'class', 'entity' or 'node'".*

In CRM Version 3.0, there are 62 Entities. A typical entity and its notation is that for "Physical Stuff" which is written as "E18 Physical Stuff" where "E18" is the unique identifier. Another example would be "E7 Activity".

Links: *"anything that may be called "attribute", "reference", "link" or "property"."*

In CRM, links are represented as "properties"; there are 107 Properties in CRM Version 3.0. Each link is represented by its forward and backward name. For example, "E7 Activity *is identified by (identifies)*: Appellation", the backward way would be written "Appellation *identifies* E7 Activity". Note the use of the colon after the property statement. Each property is numbered.

Links of links: *"are given in an indented position in parenthesis under the respective link."* This is the case where a property is associated with another property. This is a form of sub-typing. It occurs when there's a wish to provide additional information associated with link between two entities. For example, "E7 Activity '*p14 carried out by'* E39 Actor ' *in the role of* 'E55 Type"

Note. Properties of the type, 'in the role of'' were not uniquely numbered when CRM Version 3 was being developed. (This is discussed further in Chapter 4)

Superclass and Subclass: *"this relation refers to "isA" relations, "parent class – derived class", "generalisation – specialisation", etc."*

Inherited: *"links are strictly inherited to subclasses; this applies to the entities in both side of the link."*

Any instance of a subclass can inherit links, and any instance link can reference a subclass of an entity to which it points.

Figure 8 (Doerr and Stead, 2002) illustrates the above CRM terms and the "isA" relationship between each entity. Some of the entities in this diagram have been extracted from the previous hierarchy diagram set out in Figure 7. These entities are highlighted in different colours, which represent the hierarchy position of the entities in CRM. "$E_{53}, E_2, E_{52}, E_{77}$" are in the second level of the CRM hierarchy, and they all belong to the subclasses of E_1. E_{61} is in the same level with the above entities, but E_{61} is not a subclass of E_1. One entity is able to link with another entity in the lower level or entity at the same level, such as in Figure 8, E_{52} links with E_{49} at level 4, and also link with E_{61} at the same level.

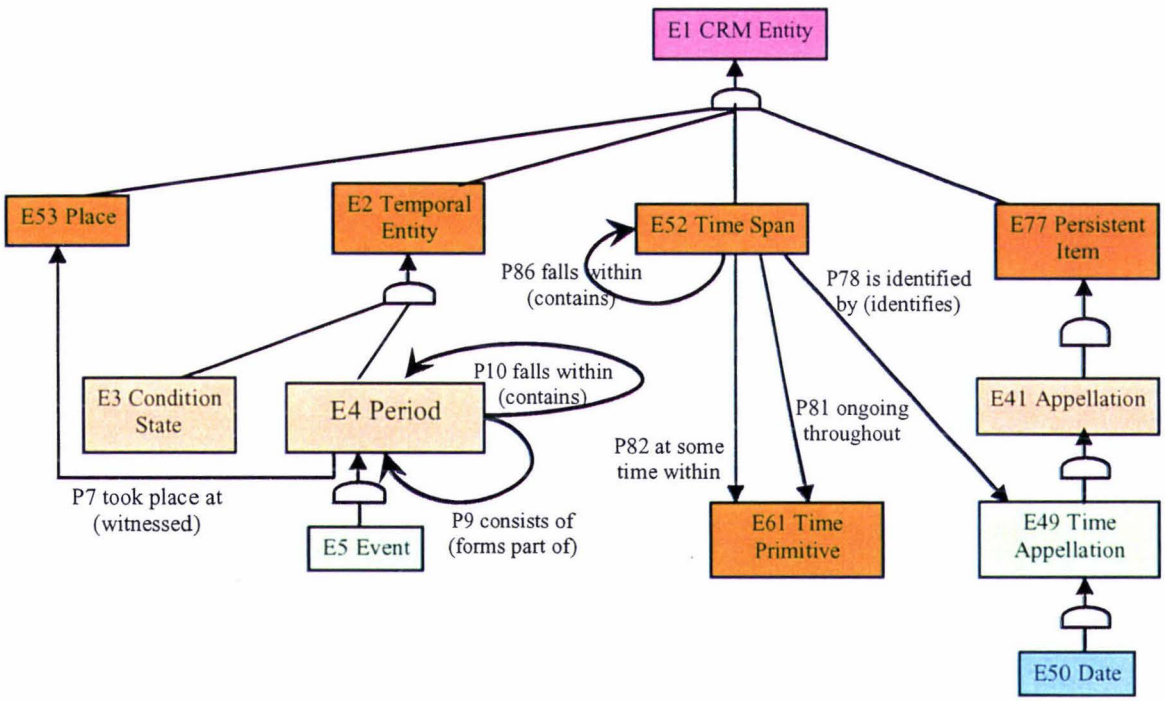


Figure 8 Relationships of some CRM Entities -- based on Doerr and Stead (2002)

In this diagram, different colours represent different hierarchical position of the entities.

3. CRM examples

Several examples are discussed in this section that show how the entitites in the CRM link to each other using various properties.

Example 1: Figure 9 is an extended version of the diagram presented by Doerr and Crofts (1999). They used the diagram to illustrate how information about the entity 'Condition Assessment' is linked to other entities, in this case, Activity, Physical

Object and Condition State. Other entities are also included (Things have Time Span, Period, Event, Physical Entity, Place) to show how properties are further extended to other entities (The figure is based on CRM Version 1).

Class Hierarchy

The classes and their subclasses are specified as follows (refer to the green boxes in Figure 9)

- 'Period' is a subclass of 'Things have Time Span', as is 'Condition State'
- 'Event' is a subclass of 'Period'
- 'Activity' is a subclass of 'Event'
- 'Condition Assessment' is a subclass of 'Activity'.
- 'Physical Object' is a subclass of 'Physical Entity'.

Inherited Properties

Properties are inherited from the parent class. In addition properties can be created within a class. In this diagram:

- 'Place' attribute in class 'Activity' is inherited from class 'Period' via class 'Event'.
- 'Time Span' attribute in class 'Condition State' is inherited from class 'Things have Time Span'.

Entities Linked by Properties

- 'Actor': linked to 'Activity' by the property: 'carried out by'.
- 'Place': linked to 'Activity' by the inherited property: 'took place at'.
- 'Physical Object': linked to 'Condition Assessment' by the property: 'assessed by'.
- 'Condition Assessment': linked to 'Condition State' by: 'has identified'.
- 'Physical Object': linked to 'Condition State' by: 'has condition'.

On this basis any 'Condition Assessment' can be dated and located.

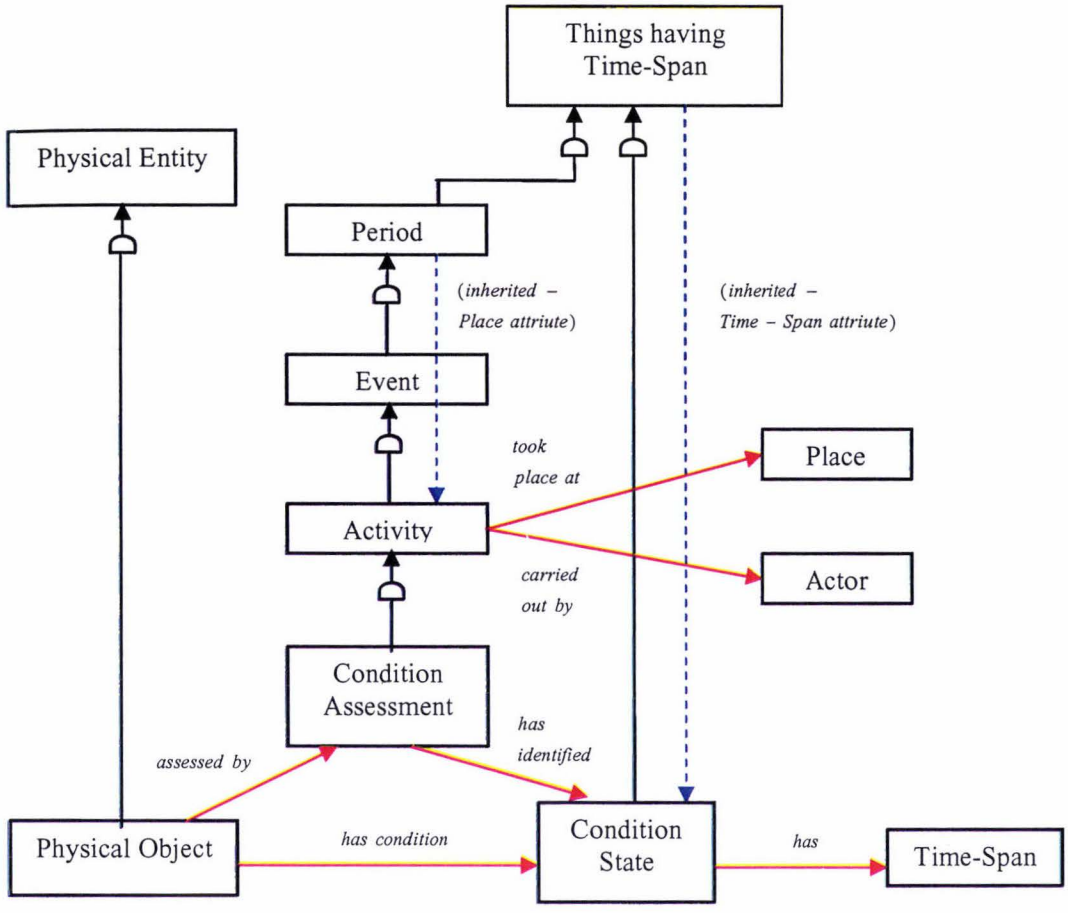


Figure 9 Relationship of Entity Condition Assessment -- Based on Doerr and Crofts (1999)

Example 2: The second example in Figure 10 illustrates how the information relating to the "Treaty of Waitangi" might be contained. All the entities and links appear in Figure 10 are based on CRM Version 3.0.

This example is interesting in that it is typical of archived historical information where 'expert' have differing views of the historical event. In the Treaty, "Captain William Hobson" represented the British Crown, and the "Maori Chiefs" represented the Maori.

For a better understanding of the inherited relationship in Figure 10 one might refer to Figure 7, which shows the hierarchical structure of the related entities, and also illustrates how links are inherited.

- 'Creation' is the class of creation of the immaterial product, such as text, music, image, law etc. It is the subclass of 'Activity'.
- 'Activity' is a subclass of 'Event', which is a subclass of 'Period', from which it inherits 'Place' attribute.
- 'Period' is subclass of 'Temporal Entity', from which it inherits the 'Time-Span' attribute.

Therefore, any 'Creation' can be dated and located.

In Figure 10, 'information' is directly associated with entities (class instances) and these are linked using properties, some of which are inherited.

- 'Creation' linked to: 'Activity' (Waitangi Meeting) by 'had specific purpose to'
- 'Creation' linked to: 'Time-Span' (6 February 1840) by 'has time-span'
- Activity' inherits the attribute 'participate in' from class 'Event' and 'Creation' inherits the attribute 'carried out by' from entity 'Activity'; therefore,
- 'Actor' (Captain William Hobson) 'participate in' 'Activity'
- 'Creation' 'is carried out by' 'Actor' (Captain William Hobson).

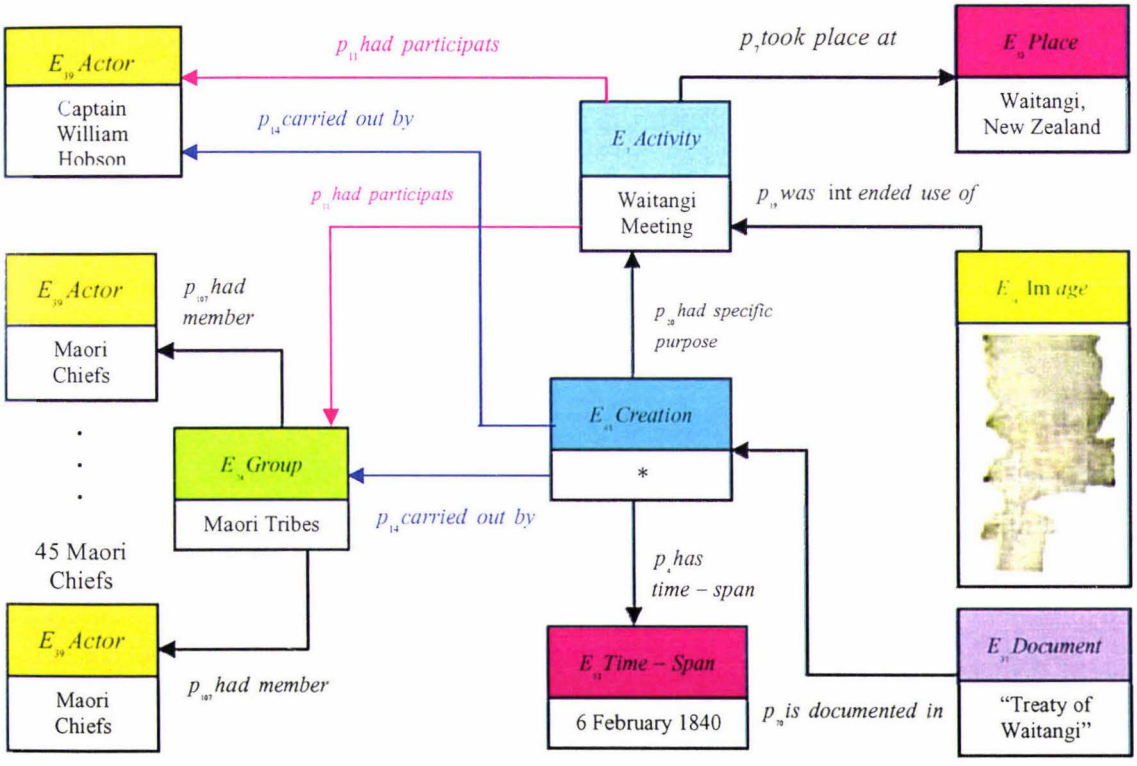


Figure 10 Mapping of "Treaty of Waitangi" to CRM

Different colours applying in Figure 10 represent different hierarchical levels of the entities in CRM

In Figure 10, there are no direct linkages from 'Activity' to 'Image', and it is same with 'Creation' and 'Document'. The following diagram (Figure 11) shows the actual path of the above two connections.

In Figure 11(a), 'Activity' is the upper-class entity of 'Creation', 'Creation' links with 'Activity' by the property p_{20}, it also inherits property p_{19} from 'Activity'; 'Creation' then connects with 'Conceptual Object' by p_{94}; and also, 'Conceptual Object' is the upper-class of 'Image', where 'Image' inherits p_{19} from, because of the connection of property p_{19}, 'Image' is able to link with 'Activity' (indicates using the red dotted line).

In Figure 11(b), 'Document' links with 'CRM Entity' by the property of p_{70}; 'Conceptual Object' inherits the property p_{70} from the connection with 'CRM Entity'; because of the connection between 'Creation' and 'Conceptual Object', 'Creation' is able to link with 'Document' by the property p_{70} (indicated using the red dotted line).

In Figure 12, the inherited relationship of the properties is illustrated and analysed in this diagram.

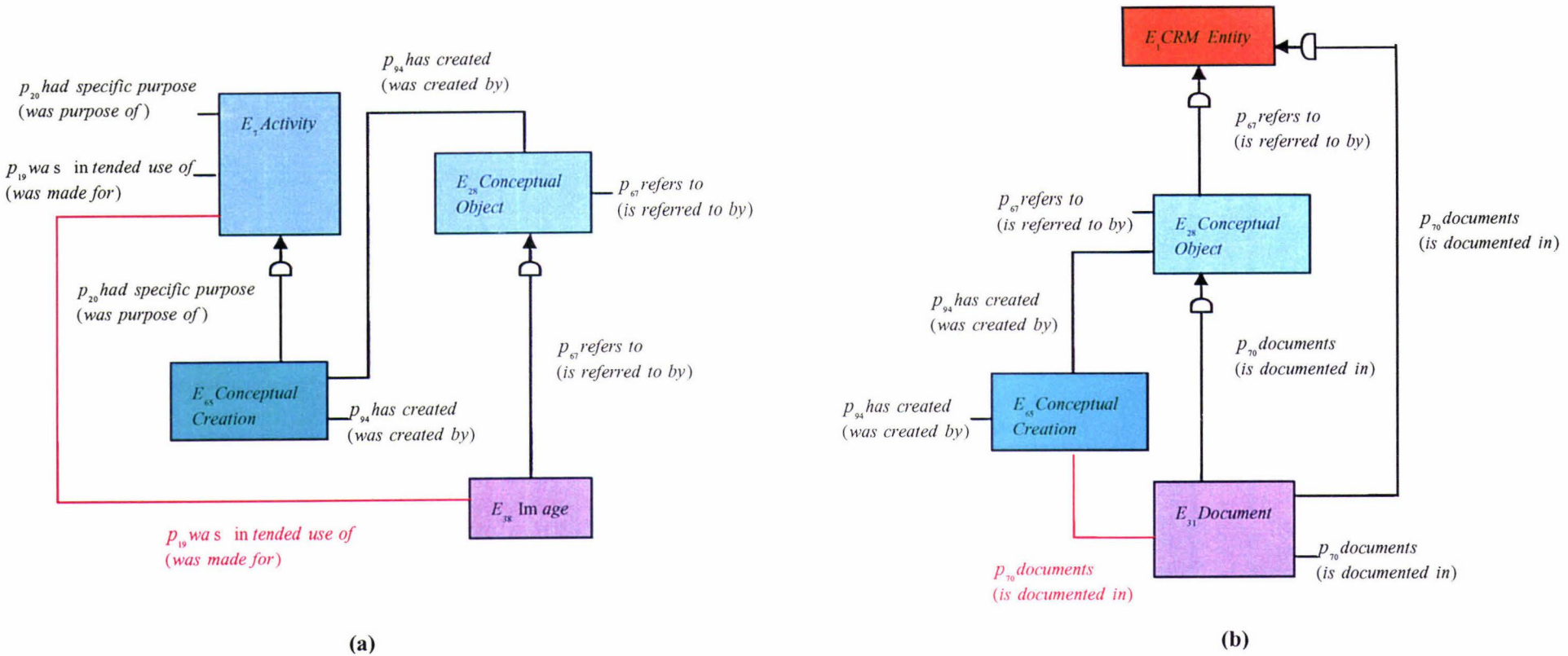


Figure 11 Actual Path of connections of 'Activity' to 'Image' and 'Creation' to 'Document'

In Figure 11, The symbol "  " is used to represent hierarchical relationship between super-class entities and sub-class entities; the black line represents the non-physical relationship between two entities which links with each other; the red line represents the connections of 'Activity' to 'Image' and 'Creation' to 'Document', each hierarchical level is shown by applying different colours, and the properties attached to the individual entities indicate where the links were declared.

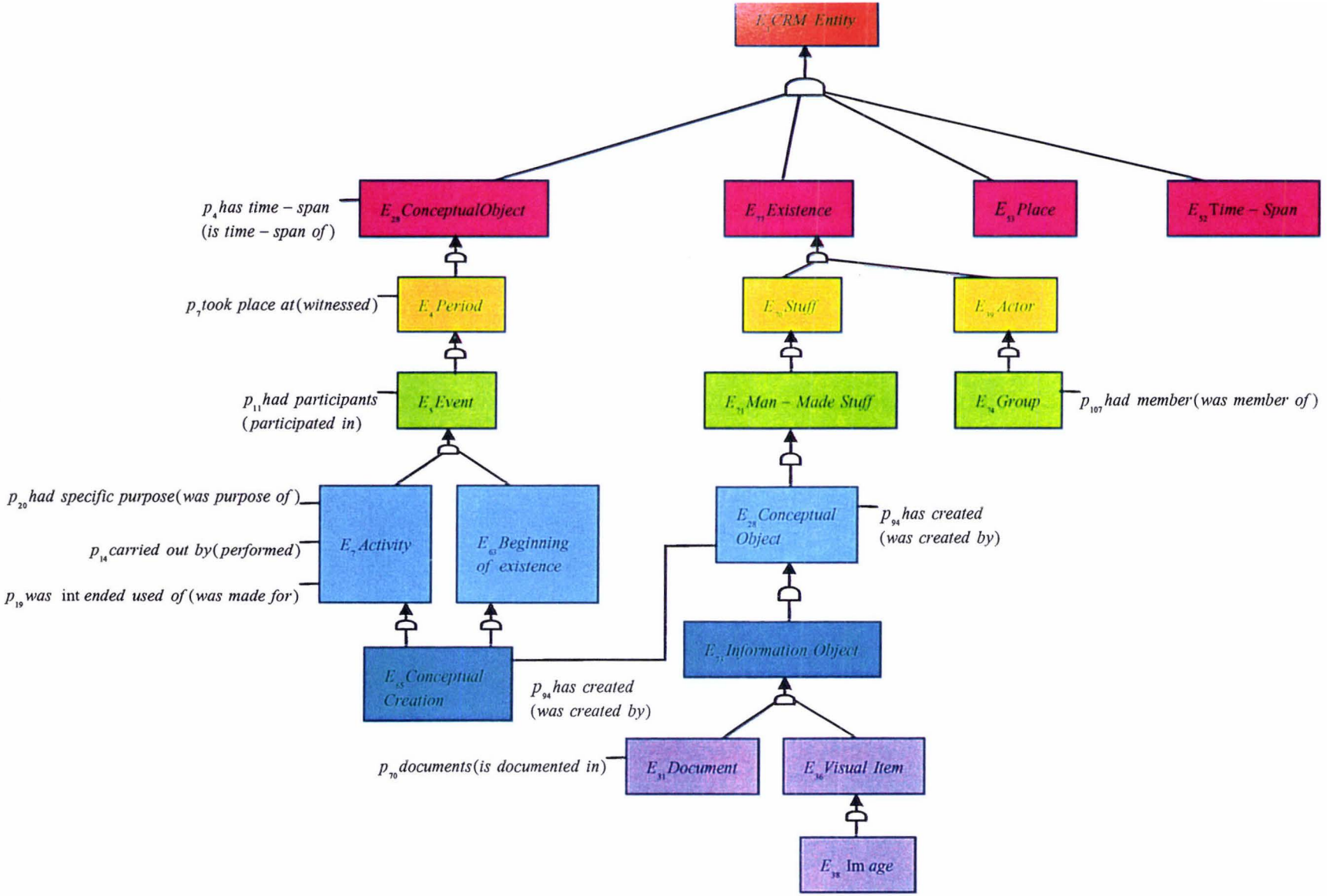


Figure 12 Hierarchical Structure of Related Entities in the mapping of "Treaty of Waitangi"

In Figure 12, the symbol "⅄" is used to represent hierarchical relationship between super-class entities and sub-class entities; the black line represents the non-physical relationship between two entities; each hierarchical level is shown by applying different colours, and the properties attached to the individual entities indicate where the links were declared.

Doerr and Crofts (1999) claim that the structure of CRM allows for a greater amount of information to be recorded. Doerr and Crofts (1999) also point out that one of the advantage of using CRM is that it allows for multiple interpretations by attributing acts of condition assessment to particular actors.

3.4.2.2 The CRM Design Process

In this section, a description of the design process used by Doerr to create the CRM is provided. Doerr (2001b) claims that the design process has been successful and he believes that this success is partly due to the emphasis being placed on the properties (links) rather than the classes (entities). This is the opposite to the way most object-oriented design methodologies are performed.

The following incremental steps illustrate the design process (Doerr, 2001b):

- **Step 1: create the list of properties of an initial set of classes.** Doerr (2001b) explains that these initial sets of classes can be made from the source model or collections of "basic level" classed on an intuitive basis. For example, the entity 'Activity (IS department function)' exists in the source model, from this entity, the property of 'carried out by: department staff' may generated as the link to 'Activity'.
- **Step 2: detect new classes from attribute values.** Properties can then be used to describe many classes. The designer may find it difficult to accept such instances as: 'table', 'horse', 'Peter Jackson' in the same category, such as the entity "Thing". In such situations the 'Thing' entity might need to be divided into sub-classes, for example 'Physical Object', 'Conceptual Object', 'Animal' or 'Person'. Doerr (2001b) suggests going to back to step 1 to redefine the properties of the newly formed classes.
- **Step 3: detect entities hidden in attributes.** Doerr (2001b) states some entities may be hidden in the already identified classes. Such as the concept of "birth" is hidden in the properties of "birth_place" and "birth_date". Doerr (2001b) suggests going back to Step 1 to describe the properties of the additional new created classes if it is necessary.
- **Step 4: property consistency test.** Doerr (2001b) suggests that the graphic presentation of the established classes and properties is useful for consistency control. It is important to ensure that a property can be expressed clearly and unambiguously when read from the domain class to the range class and vice

versa. Inability to achieve this goal might suggest the need to create new properties or even rearrange the domain and range. If necessary, one might need to go back to Step 2 to detect any hidden properties.

- **Step 5: create the class hierarchy.** When a new class is generated, one needs to determine where the new entities should be placed in the new structure. Conflict might arise between the existing properties and the properties of the new entity, and these will need to be resolved. Doerr suggests that there may be a necessity to reduce the number of the abstract classes in order to take out "overspecialized" classes and properties. He suggests going back to Step 1 to describe the additional classes or to Step 4 for merging the properties.
- **Step 6: create property hierarchies.** According to Doerr this may lead to the detection of more properties and inconsistencies. He suggests check Step 4, else end with Step 7.
- **Step 7: closing up the model.** Continually iterating between the seven steps could be a problem. In general, process will naturally stop when primitive values, such as numbers, time-span. "peripheral" properties have been declared.

No model is really complete until it has been tested (validated) using real-world data from a collection.

3.4.2.3 Versions of CRM

It has taken over ten years to develop the CIDOC CRM. Over that time CIDOC DSWS has released several versions of the CRM. The first version of CRM was completed in 1998, and its model was presented at the triennial ICOM conference in Melbourne in 1998. CRM version 2.2 was successfully submitted to ISO TC46 as new work item, and now it is currently undergoing evaluation by the International Standards Organisation as Committee Draft ISO/CD 21127 and the CIDOC CRM Special Interest Group to become an ISO standard (ICOM/CIDOC Documentation Standards Group, 2003).

The CRM structure diagram presented in the previous section is based on Version 3.3.2. This researcher chose this version, as it was the most recent one when she started working on the CRM structure in earlier 2003.

In Chapter 5, version 3.0 was used to replicate the validation of the EAD, Dublin Core and AMICO collection systems. To maintain consistency and comparability, the validation of the two New Zealand centres also made us of version 3.0.

Note: Version 3.4.9 (Crofts, et al, 2003) is the latest version of CRM, which was published at the end of November 2003. According to ICOM/CIDOC Documentation Standard Group, the current version has covered the intended scope of CRM outlined in July 2001, and the general functionality required by the group has been successfully fulfilled, so that further improvement of the model will focus on improving the clarity of the text.

Two improvements to the documentation have been introduced since the completion of this work: the need to number the "property builds on property" and the cardinality of the relationships. Both of these are introduced independently by the research prior to the latest version of the CRM being published.

4 Analysis 1: Diagrammatical and Notational Representation

4.1 Introduction

One of the aims of this thesis is to understand and explain the CRM approach to expressing cultural knowledge within a semantic framework. Two key techniques used to document the mapping process are adopted in this report: diagrams and mathematical notation. The researcher is of the opinion that these techniques have unique properties in that they enhance understanding and communication of ideas, and provide a means for investigators to explore a proof or concept. The following sections in this chapter provide some justification for using these forms of representation.

4.2 Diagrammatical representation

4.2.1 Introduction

Diagrams or pictures are some of the oldest forms of human communication. Their use is not limited to representation but can also be used to carry out certain types of reasoning. Such a capacity makes a diagram a useful tool for mathematicians, logicians and computer scientists. Diagrams are usually adopted as a heuristic tool to explore a proof or concept (Newsham, 1995; Engelhardt, 1998), however, diagrammatic systems are currently used in a wide area of contexts, logic teaching, automatic reasoning, computer programming specification and many other situations.

4.2.2 Definition

In a general sense, a diagram is a user interface term for a representation of some group of information that makes use of structural or symbolic representation. The New Oxford Dictionary of English (1999) provides the following definition:

> *"1. Geom. A figure composed of lines, serving to illustrate a statement or to aid in a demonstration 1645.*
>
> *2. Illustrative figure giving an outline or general scheme of an object and its various parts 1619.*
>
> *3. A graphic representation of the course or results of any action or process or its variations."*

As shown in the following section diagrams have more utility than that suggested in the New English Dictionary. In this thesis, in addition to using diagrams as a means of

illustration and communication they are used to explore and validate concepts – or more specifically mapping and notation.

4.2.3 Taxonomy

The use of diagrams has been the subject of much analysis (Dale, 1969; Cox and Brna, 1995; Engelhardt, 1998) and several taxonomies have been proposed that allow users of diagrams to validate in some modest way the effectiveness of their actions. For example, Martin and McClure (1985) and Newsham (1995) examined the efficacy of the use of notation and diagrams used in software engineering. Price et al did the same for visual programming languages (Price et al, 1993). The ergonomic implications of these diagrams have been categorised in the cognitive dimensions of notations (Green and Blackwell 1998). Further examples include the selection of representations for educational contexts (Goldsmith 1984), or in cartography, typography, and graphic design (Bertin 1967, Engelhardt 1998).

- ***Nine aspects of diagrams and diagram use***

Drawing on the work of the previously mentioned researchers and others, Blackwell and Engelhardt (1998) proposed a meta-taxonomy consisting of nine aspects, which is summarised in the following table.

Table 1 Nine aspects of diagrams and diagram use (Blackwell & Engelhardt, 1998)

Signs – the components of a diagram	
1.	Basic graphic vocabulary
2	Conventional elements
3	Pictorial abstraction
Graphical structure of a diagram	
4	Graphic structure
Meaning	
5	Mode of correspondence
6	The represented information
Context related aspects	
7	Task and interaction
8	Cognitive processes
9	Social context

The nine aspects can be divided into representation-related aspects (1-6) and context-related aspects (7-9). According to Blackwell and Engelhardt this set of nine taxonomic aspects can be used to examine the format and design of a diagram.

- ***Representation-related aspects***

The representation-related aspects are based on the semiotic dyad proposed by Saussure (1966), which links a representation to its meaning.

Representation-related aspects relate either to the diagram itself (1-4), or to its meaning (5-6). These aspects are concerned with either the signs that are the components of the diagram (1-3), or with the graphic structure of the diagram (4). These aspects regarding meaning are concerned with either mode of correspondence (5) or with the represented information (6).

These 6 aspects as described by Blackwell and Engelhardt (1998) are shown in Figure 13 below.

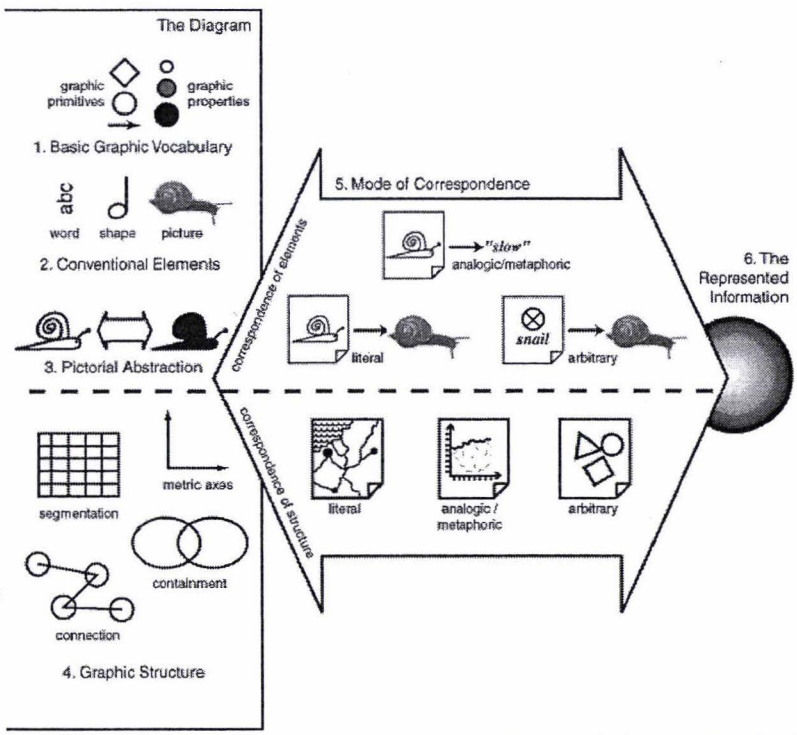


Figure 13 Representation-related taxonomic aspects (Blackwell & Engelhardt, 1998)

- ***Context-related aspects***

The intended result of the communication is referred to as the interpretant in Peirce's semiotic triangle (Peirce, 1932). However, the result of a communication is not independent of the context in which it is located. For example, the diagram may be used to communicate information to an audience such as at a lecture. This would be

more in keeping with the principle of the semiotic triangle. On the other hand, it may be used by an individual who is using the diagram to analyse of solve a problem. Blackwell and Engelhardt suggest that it would be better not to refer to an interpretant, but to a range of possible diagram contexts (Aspects 7-9, see Figure 14). This is particularly relevant in this thesis where the diagrams derive their status from the context of task and interaction (7) and where the user interacts actively with the diagram. Equally important in the case of this thesis is the stimulation of the cognitive processes (8) where the researcher may be regarded as an independent agent, with the context of diagram use being the researcher's mental state. Finally the social context (9) is also important as this report is being used to communicate information to a wider socially constructed audience.

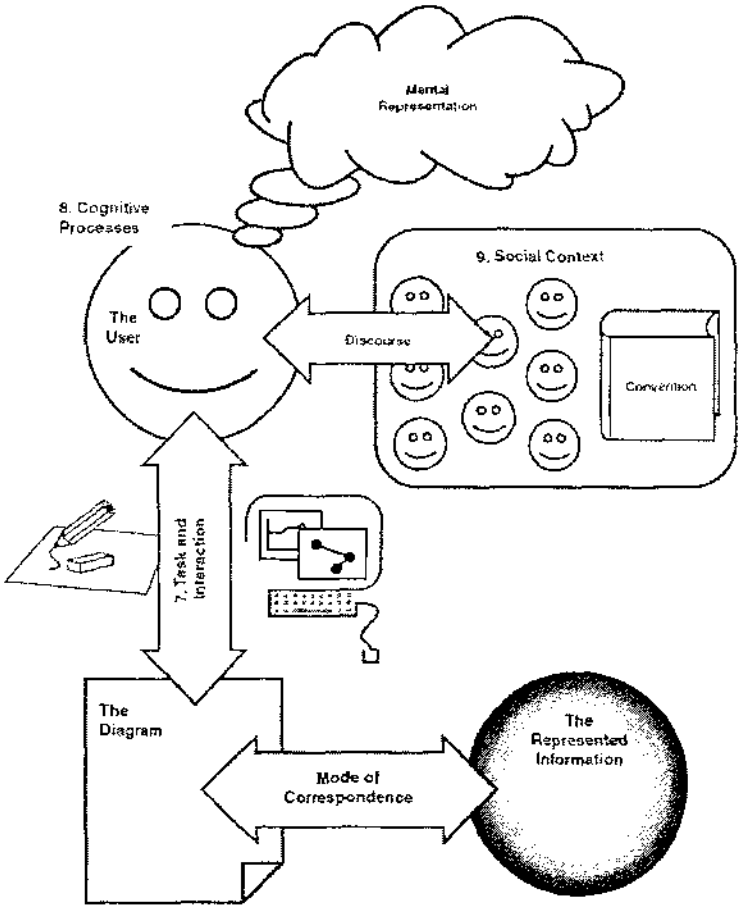


Figure 14 Contextual taxonomic aspects (Blackwell & Engelhardt, 1998)

Basic graphic vocabulary (1)

The basic graphic vocabulary consists of the graphic primitive elements together with their properties. Typical examples include "point, line, area" and "colour, size, shape".

Conventional elements (2)

Common conventional elements include words, shapes and pictures – these could be further simplified to the dichotomy "abstract vs pictorial". The simple distinction between words and pictures is included in *mode of correspondence* (5), while that between shapes and pictures is included in *pictorial abstraction* (3).

Pictorial abstraction (3)

Concerning the depiction of physical objects or scenes, a *continuum* of pictorial abstraction can be observed, from the very realistic via the schematic to the completely abstract.

Graphic structure (4)

Graphic structure is concerned with the organisational principles according to which individual signs are combined into a diagram. Typical configurations include; linear sequence, chart, table, tree structure, networks, process and mapping.

Mode of correspondence (5)

This involves the kind of relationship between a representation and its meaning. These modes could be literal, metaphorical, direct, indirect, iconic or symbolic. A sign could have different meanings in different contexts.

The represented information (6)

As one would expect information can be classified in many ways. Blackwell & Engelhardt (1998) use as an example the London Underground diagram. The position of the various tube stations is spatial information whereas the 'lines' represent ordinal information (relational) or perhaps even process information.

Task and interaction (7)

The user's interaction with a diagram can be both physical and mental. Again using the London Underground diagram as an example of task and interaction. Blackwell and Engelhardt claim that travellers frequently use their fingers to trace the route to their destination. Blackwell & Engelhardt (1998), provide further examples - the use of computational tools such as diagram parsers and editors. In this thesis the mapping diagrams are used to help the researcher generate mathematical notation which could

in turn lead to the generation of programming code to extract cultural data from databases as a precursor to web-based interoperability.

Cognitive processes (8)

The user of a diagram is the innovator and they tend to choose an approach that meets their particular cognitive needs. In this thesis the need of the researcher was to help create a mental model of the mapping process within the CRM. This approach is not unusual as the cognitive implications of diagram properties related to perception, interpretation and problem solving, as well as individual differences in ability, expertise or strategy, (Blackwell and Engelhardt, 1998).

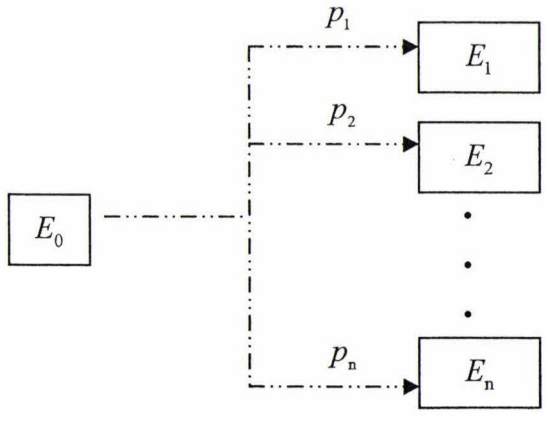
Social context (9)

The way users interpret and use diagrams depends on both the cultural context as well as the conventions of particular media types (Blackwell & Engelhardt, 1998). The user attempts to determine what information is present in the diagram and seeks ways of accommodating that information with knowledge gained from other sources. The contents and form of a diagram need to be considered in terms of its intended audience and discourse.

4.2.4 Example

Example 1: A diagram, which can be found later in this report, is used to illustrate the nine aspects of diagram structure and use. This diagram and its accompanying notation are discussed to a greater depth later in the report.

For the n pairs of objects (Entities) E_0, E_1 ; E_0, E_2 ; ... E_0, E_n, there are n morphisms represented by $p_1, p_2 ... p_n$.



Subjecting this diagram to the nine aspects of diagrams and diagram use is set out in the following table:

1. Basic graphic vocabulary	The diagram uses two implantations "rectangle" (entities) and "lines" (connections).
2. Conventional elements	Both "schematic" elements (rectangles, lines, arrows) and "verbal' elements (words – entity, properties).
3. Pictorial abstraction	None figurative.
4. Graphic structure	There is significant number of structural elements. "Organisation" is expressed by linking entities. In fact, the diagram can be viewed as a structure, a process diagram or even a network chart.
5. Mode of correspondence	The diagram is "symbolic" in the use of a range of arrowed lines and rectangles.
6. Represented information	The use of arrows, logical sequencing of actions and colour indicate both spatial and ordinal information.
7. Task and interaction	The user of the diagram, either the creator or the reader would trace the path linking the entities in order to gain a better understanding of the mapping processes. It is quite likely that a user would use a pen or other pointing tool to trace the links.
8. Cognitive processes	The diagram uses different forms of lines (solid, broken), colour discrimination, font style and size as pre-requisites of the diagram's usability.
9. Social context	The diagram meets the needs of the researcher, the supervisor and other members of the research community. Other readers of this report would invariably view it from a different perspective.

Example 2: Consider the set of entities and properties in Figure 15. An apparent issue of say E_{65}*Conceptual Creation* and E_{31}*Document* where at first sight there appears to be no formal property linking the two (Figure 15). In Figure 15, no links can be found between the entities E_{28} and E_{31}, however, because of the connection of: $E_{28} \rightarrow E_1 \rightarrow E_{28}$, E_{28} is able to inherited the property p_{70}*documents* (*is documented in*) from E_{31}; also, because of the connection between E_{28} and E_{65}, E_{65} is able to inherit the property p_{70}*documents* (*is documented in*) from E_{28}, that's how the connection between E_{28} and E_{31} is constructed. The concept of property inheritance is raised in section 4.3.4. This inheritance properly could have been indicated in the diagrams using say, "$\overrightarrow{p_{70}}$" but was thought unnecessary. It is important to point out that some

information in the CRM has deliberately not been developed into formal properties or links (Crofts, et al, 2001, p. iv).

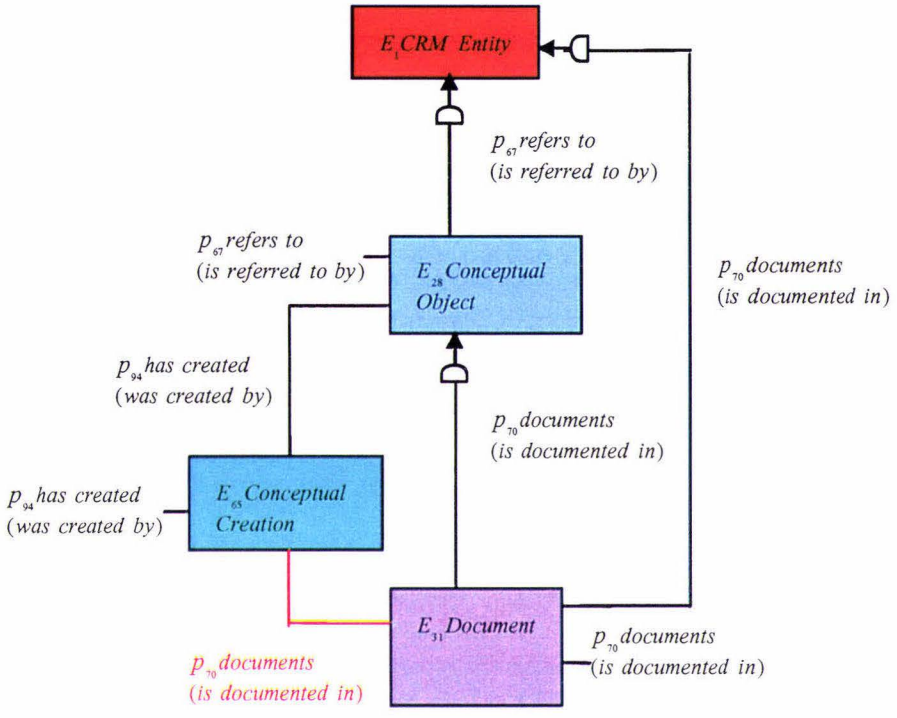


Figure 15 Example of non-physical relationship connects by inherited link

In this diagram, different colours represent different hierarchical position of the entities. The symbol "  " is used to represent the hierarchical relationship between the entities; the solid black line represent the connection between two entities; and the red solid line represent the connection between two non-physical entities

4.2.5 Summary

The importance and value of using diagrams to enhance understanding of a complex environment has been explored and has laid the foundation for the analysis of the CRM semantic model and its validation against both international and national sources of cultural artefacts. The use of diagrams will not only enable the underlying ideas of the CRM to be explored but will, perhaps, enable some of the mapping issues to be identified. The diagrams have led to the development of a mathematical notation as shown in the following section within this chapter and the graphical notation successfully addresses all the requirements of the CRM.

4.3 Mathematical representation

4.3.1 Introduction

As mentioned at the beginning of this chapter one of the aims of this research is to understand and explain the CRM approach for expressing cultural knowledge within a semantic framework. The development of a mathematical notation is considered to be an integral part of that process.

The CRM consists of a set of entities $\{E_1, E_2, ...E_n\}$. Some of these entities are linked to each other by means of the property p_i. This relationship can be represented by $(p_i)E_j :\rightarrow E_k$. However, in the CRM once the pair p_i and the source entity E_j have been identified the destination entity E_k is redundant.

The following is an extract from Appendix 2 showing the mapping of the entity E_{18} to E_{54}, E_3, E_{57} and E_{18} using the properties p_{43}, p_{44}, p_{45} and p_{46}.

P43	Physical Stuff	E18	has dimension (is dimension of)	Dimension	E54
P44	Physical Stuff	E18	has condition (condition of)	Condition State	E3
P45	Physical Stuff	E18	consists of (is incorporated in)	Material	E57
P46	Physical Stuff	E18	is composed of (forms part of)	Physical Stuff	E18

Given the above observations, the application of category theory to the CRM was seen to be a real possibility for developing an acceptable mathematical notation. As shown above, what is of interest in the CRM is not the individual objects (entities) but the morphisms (mapping - properties) between the objects. This is the very essence of category theory too.

4.3.2 Definition, Basic Concept and Syntax

Marquis (2003) described category theory as "a generalised mathematical theory of structures and systems of structures." Taking a more formal perspective we have:

> Let category **C** be described as a collection of objects O, where the objects of **C** satisfy the following conditions:
>
> - For every pair of objects a, b there is a collection $\mathbf{M}_{(a,b)}$, mapping from a to b in **C**, this can be written as, f: $a \rightarrow b$
> - For every triple of objects a, b, c, there are two partial operations called the composition of morphisms (mapping) - $\mathbf{M}_{(a,b)}$ x $\mathbf{M}_{(b,c)}$. These can be

written as, (**f** o **g**): $a \rightarrow c$ where f: $a \rightarrow b$ and **g**: $(b \rightarrow c)$. Or more succinctly, **gf**: $a \rightarrow c$

- For every object a there is an identity morphism $\mathbf{Id}_a$ in $M_{(a,a)}$

Two axioms need to be satisfied:

- Associativity:

  if **f**: $a \rightarrow b$, **g**: $(b \rightarrow c)$ and **h**: $(c \rightarrow d)$ then **h** o (**g** o **f**) = (**h** o **g**) o **f**

- Identity:

  if **f**: $a \rightarrow b$, then ($\mathbf{Id}_b$ o f) = **f** and (**f** o $\mathbf{Id}_a$)

All the above conditions and axioms are met within the CIDOC CRM.

4.3.3 Examples

In developing this notation and its subsequent application of the notation within this thesis the following strategy was adopted.

1. In the first instance a diagram was drawn showing the relevant mapping.
2. Mathematical expression was subsequently created using the diagram as a guiding tool.

It is expected as the user becomes better acquainted with the notation, there will be less reliance on the diagram.

Diagram notation

The following notation was adopted:

- A solid line represents a required morphism ——————
- A interrupted dotted one and only one — · — · —
- A dotted line represents zero, one or more .....

Category Theory notation

The following notation was adopted mirroring the three diagram notations above.

- Required morphism ——————

  $< p_1.p_2...p_n >: E_o \rightarrow \prod_1^n E_i$

- One and only one — · — · —

$$[p_1.p_2...p_n]: E_o \rightarrow \coprod_{i=1}^{n} E_i$$

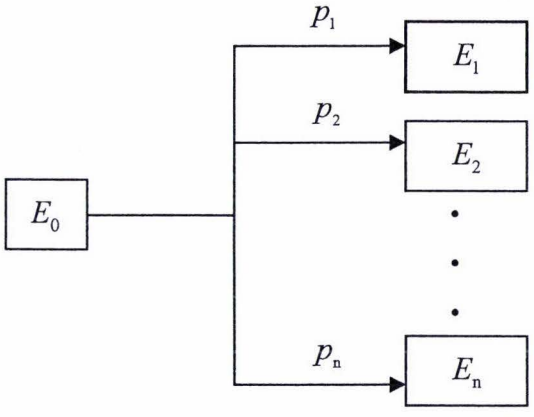
- Zero, one or more .....

$$[\![p_1, p_2, ..p_n]\!]: E_0 \rightarrow \bigsqcup_{i=1}^{n} E_i$$

The third form of notation is not part of 'standard' category theory. Nevertheless it does not seem an unreasonable mapping expression.

Example 1: Equivalent to the Product (Compulsory relationship)

- Consider the following diagram. For the n pairs of objects (Entities) E_0, E_1; E_0, E_2; ... ; E_0, E_n there are n morphisms represented by $p_1, p_2, ..., p_n$.



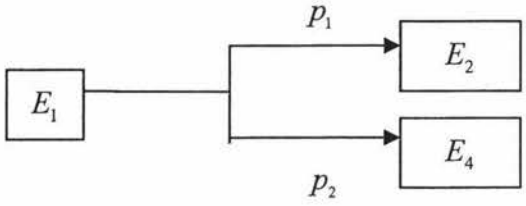
Since each and every morphism is compulsory then we have the equivalent of a Product relationship. This can be represented in category theory by the notation (using angle brackets and a $\prod$):

$$\langle p_1.p_2...p_n \rangle : E_o \rightarrow \prod_{i=1}^{n} E_i$$

In CRM, as with category theory, once the source entity and its associated link are known then the destination entity is known. Removing this redundancy allows the expression to be written:

$$\langle p_1.p_2...p_n \rangle : E_o$$

- Consider the following specific example. In this diagram, E_1 relates with both E_2 and E_4 using the properties p_1 and p_2:

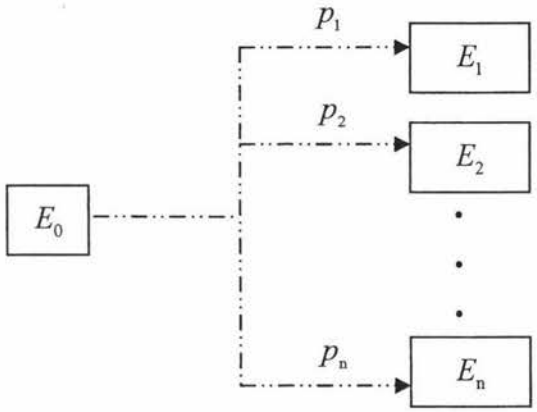


This can be expressed as: $\langle p_1,p_2\rangle : E_1 \rightarrow E_2 \times E_4$ or $\langle p_1,p_2\rangle : E_1 \rightarrow \prod(E_2, E_4)$

After eliminating the redundant component: $\langle p_1,p_2\rangle : E_1$

Example 2: Equivalent to the Co-product (one and only one relationship)

a Consider the following diagram. For the n pairs of objects (Entities) E_0 , E_1 ; E_0 , E_2 ;…; E_0 , E_n there are n morphisms represented by $p_1, p_2, \ldots, p_n$.



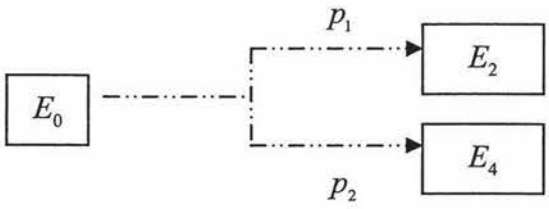
In this situation one and only one pair of objects (Entities) is mapped. This is equivalent to the coproduct relationship. This can be represented in category theory by the notation (using square brackets and an upside down $\prod$):

$$[p_1, p_2, \ldots, p_n] : E_o \rightarrow \coprod_{i=1}^{n} E_i$$

Removing the redundant components:

$$[p_1, p_2, \ldots, p_n] : E_o$$

b Consider the following specific example. In this diagram, E_1 relates with one and only one of E_2 and E_4 using the properties p_1 and p_2 :

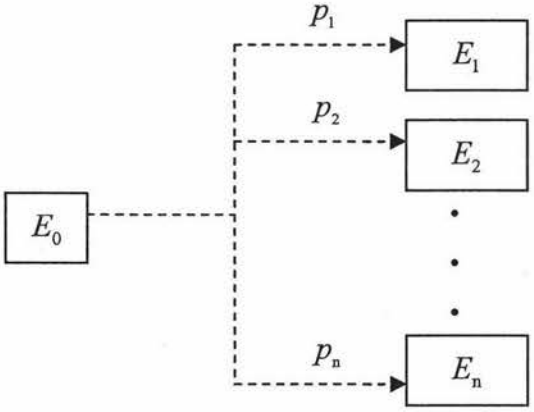


Notation for the diagram can be written as: $[p_1, p_2]: E_0 \rightarrow E_2 \oplus E_4$ or $[p_1, p_2]: E_0 \rightarrow \amalg(E_2, E_4)$

Removing the redundant part, the notion becomes: $[p_1, p_2]: E_0$

Example 3: No equivalent – new notation required (Optional relationship)

a Consider the following diagram. For the n pairs of objects (Entities): E_0, E_1 ; E_0, E_2 ; ... E_0, E_n there are n morphisms represented by $p_1, p_2, ..., p_n$. E_0 is able to link zero, one or more of the entities.



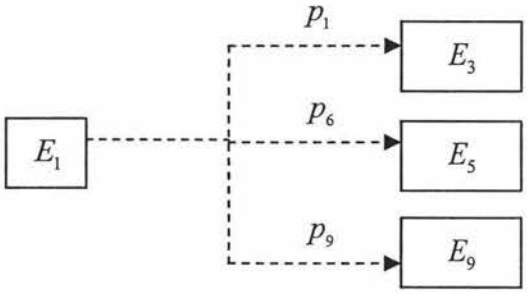
This can be represented using the proposed new notation (using double brackets and a square):

$$[\![p_1, p_2, .. p_n]\!]: E_0 \rightarrow \overset{n}{\underset{i=1}{\square}} E_i$$

Removing the redundant components:

$$[\![p_1, p_2, .. p_n]\!]: E_0$$

b Consider the following specific example. In this diagram, E_1 relates with zero, one or more of E_3 , E_5 and E_9 using the properties p_2 , p_6 and p_9 .



This can be represented using the new notation:

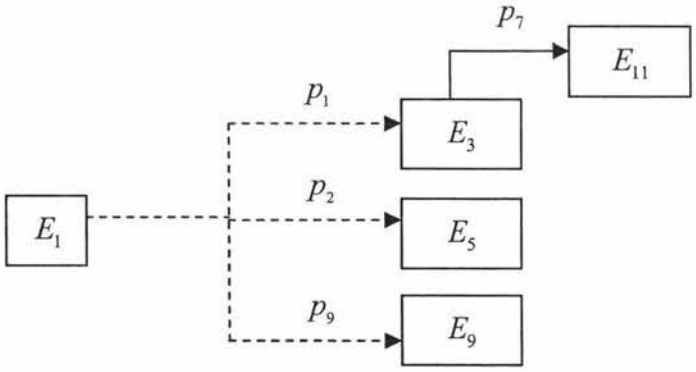
$[\![p_2, p_6, p_9]\!] : E_1 \rightarrow \square(E_3, E_5, E_9)$

Removing redundant components:

$[\![p_2, p_6, p_9]\!] : E_1$

Example 4: Combination of Compulsory and Optional

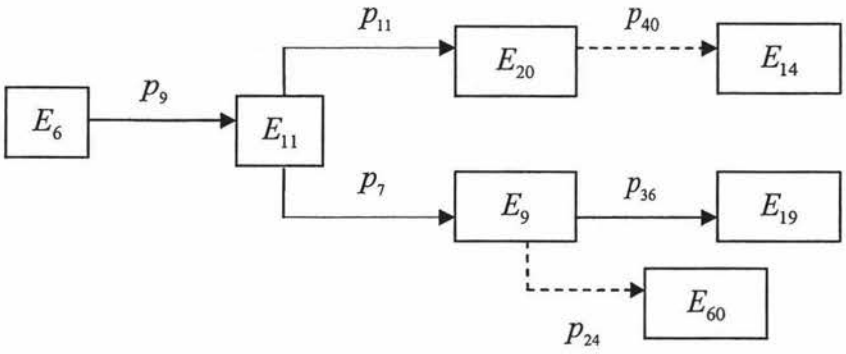
a. Consider the following diagram. The first part of the diagram is the same as Example 3. The second part, E_3 relates with E_{11} using the property p_7. The latter mapping is compulsory if and only if p_1 is present.



In the situation, the relationship of $E_1 \rightarrow E_3 \rightarrow E_{11}$ is very similar with the composition of morphisms: $\mathbf{M}_{(a,b)}$ x $\mathbf{M}_{(b,c)}$. So $E_1 \rightarrow E_3 \rightarrow E_{11}$ can be expressed using the formula: $\langle p_7 \rangle p_1 \rightarrow E_{11}$. When combine this with rest of the diagram, the whole diagram can be represented as: $[\![\langle p_7 \rangle p_1, p_2, p_9]\!] : E_1 \rightarrow \square(E_3 \times E_{11}, E_5, E_9)$.

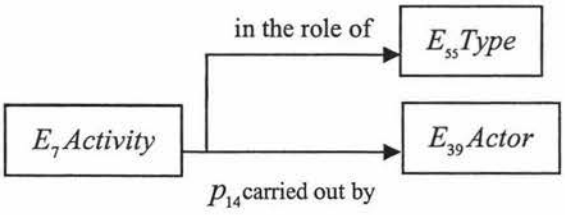
b. Consider the following diagram. The diagram can be treated in two parts: $E_6 \rightarrow E_{11} \rightarrow E_{20} \rightarrow E_{14}$ and $E_6 \rightarrow E_{11} \rightarrow E_9 \rightarrow (E_{19}, E_{60})$. The first part can be expressed as: $\langle [\![p_{40}]\!] p_{11} \rangle p_9 : E_6 \rightarrow E_{20} \oplus (E_{20} \times E_{14})$; the second part can be expressed as: $\langle p_{36}, [\![p_{24}]\!] \rangle p_7 p_9 : E_6 \rightarrow E_9 \oplus (E_{19} \times E_{60})$. When combining the two parts together, the whole diagram can be expressed as:

$$\langle [\![p_{40}]\!] p_{11}, p_7 \langle p_{36}, [\![p_{24}]\!] \rangle \rangle p_9 : E_6 \rightarrow (E_{20} \oplus (E_{20} \times E_{14})) \times (E_9 \oplus (E_{19} \times E_{60}))$$



Example 5: Property builds on property

Property builds on property describes a property (link) attached to another property as shown in the following diagram. See the example below. In this diagram, $E_7 Activity$ is carried out by $E_{39} Actor$, and the link "in the role of" is used to describe the role of the Actor plays in this particular activity.



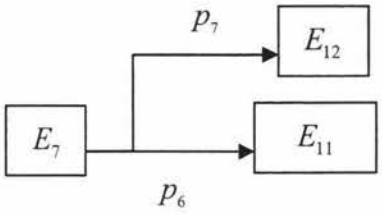
The use of an unnamed property 'in the role of' allows the nature of the Actor's participation to be specified, thus increasing the richness of the information. To express this situation using the new notation requires an extension to the set of CRM properties, some of which are displayed in the following table.

Link ID	Link Name	Property Name Links to
P108	in the role of: Type	carried out by (performed)
P109	mode of depiction: Type	depicts concept (is depicted by)
		depicts event (is depicted by)
		depicts object (is depicted by)
P111	has type: Type	has note
		right held by (has right on)
		refers to (is referred to by)
		has title (is title of)
P112	has note: String	right held by (has right on)
P113	mode of use: String	used object (was used for)

Note: The problem of unnamed properties was addressed in later versions of the CRM, which were published about the time this thesis was completed. For example, $p_{14.1}$ was used for 'in the role of', $p_{62.1}$*mode of depiction*, $p_{3.1}$*has type* and $p_{67.1}$*has type*. In the following mapping for the international and New Zealand validation for CRM, numbers used in the latest version 3.4.9 of CRM is applied.

Incorporating the 'property builds on property' to the new category theory is the subject of the following four situations:

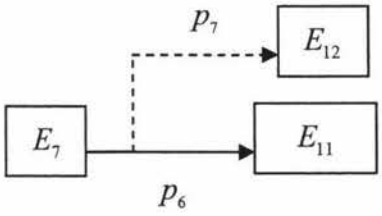
a. Consider the diagram, E_7 relates to E_{11} using the property p_6. In this case, the relationship between E_{12} and p_6 is compulsory, as well as the link p_7 to E_{12}. E_7 is able to link with p_6 through the combination of E_{11}, E_{12}



The diagram can be expressed as: $\langle p_6, p_7 \rangle : E_7 \rightarrow E_{11} \times E_{12}$

After eliminate the redundant part, the formula becomes: $\langle p_6, p_7 \rangle : E_7$

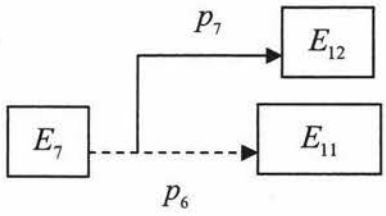
b. Consider the diagram, the link between E_{11} and p_6 is compulsory, however, the link to E_{12} through p_7 is optional. E_7 is able to link with p_6 in two ways: E_7 link with E_{11} or through the combination of E_{11} and E_{12}



The diagram can be expressed as: $[p_6, \langle p_6, p_7 \rangle] : E_7 \rightarrow E_{11} \oplus (E_{12} \times E_{11})$

After eliminate the redundant part: $[\![p_6, (p_6, p_7)]\!] : E_7$

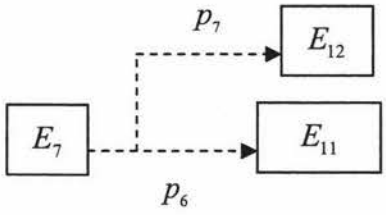
c. Consider the diagram, the link between E_{12} and p_6 is optional, however, the link to E_{12} through p_7 is compulsory. E_7 is able to link with p_6 in two ways: E_7 link with combination of E_{11}, E_{12}, or E_7 has no link with any entities.



The diagram can be expressed as: $[\![\langle p_6, p_7 \rangle]\!] : E_7 \rightarrow \square\{E_{12} \times E_{11}\}$

After eliminate the redundant part: $[\![\langle p_6, p_7 \rangle]\!] : E_7$

d. Consider the diagram,the link between E_{12} and p_6 is optional and the link to E_{12} through p_7 is optional as well. E_7 is able to link with p_6 in three ways: E_7 link through the combination of E_{11}, E_{12} or E_7 link with E_{11} or E_7 has no link with any entities.



The diagram can be expressed as: $[\![p_6, \langle p_6, p_7 \rangle]\!] : E_7 \rightarrow \square\{E_{11}, E_{12} \times E_{11}\}$

After eliminate the redundant part: $[\![\langle p_6, p_7 \rangle]\!] : E_7$

Note: the cardinality relationship presented here has been solved in the latest version of CRM (version 3.4.9) which were published about the time this thesis was completed. Therefore, the cardinality relationship still covered in this thesis.

4.3.4 Summary

The possibility that category theory notation could be used to represent mapping within the CRM appears quite likely on the basis of this preliminary work. The two

forms of representation will be applied in the following chapters to some of the validation activities undertaken in support of the CRM approach.

It is expected that some issues will arise with the notation and possibly some issues with the CRM model itself.

There are likely to be a number of possible advantages of representing mapping using category theory the major one being the possibility of facilitating the parsing of data stored within a database to a web-based language for presentation over the Internet. However, the implementation or even research into that field of endeavour is outside the scope of this report.

Category notation appears to meet all the requirements of the CRM except, perhaps, the case where properties have properties, which may link to other entities. Typically these properties are used for dynamically modified properties such as roles. As shown above, a work around has been carried out but another approach might be necessary.

Note:
The lack of a formal property mentioned in Section 4.2.5 does not adversely affect the use of Category Theory.
Using the example in Section 4.2.5, there is: $< \overline{p70} >: E65 \rightarrow E31$ where $\overline{p70}$ is inherited from the $E_1 CRM\ Entity$.

The above expression could be reduced to: $< \overline{p70} >: E65$ as the destination entity $E65$ is uniquely determined by the source entity and the property as one would expect in Category Theory.

5 Analysis 2: International Validation of the CRM

Two of the aims of the research are covered within this chapter. Aim 1: To gain a better understanding of the CRM and Aim 3: To apply the new tools to replicate international validation of the CRM. (Refer to Section 1.6)

The following three publications form the prime sources for this data:

- EAD: Mapping of Encoded Archival Description to the CRM version 3.0 (Theodoridou and Doerr, 2001)
- DC: Mapping of Dublin Core to the CRM version 3.0 (Doerr, 2000)
- AMICO: Mapping of the AMICO data dictionary to the CRM version 3.0 (Doerr, 2001a)

The reason these three validation exercises were adopted is that they represent the spectrum of activities performed by museums and art galleries. The EAD is used for describing the content of bibliographic material, AMICO for artwork, and the Dublin Core is now being adopted as a 'standard' for retrieving and displaying data across the Internet. Doerr (2000, p2) suggested that *"These mapping are not meant to be reversible in a formal sense. Rather they describe how to transform descriptions in one structure into an equivalent description in the CRM structure with the same meaning, to the degree the contents under investigation fall under the scope of the CRM."*

5.1 Mapping of the generic elements in EAD to CRM (based on CRM Version 3.0)

Encoded Archival Description (EAD) Document Type Definition (DTD) is a Standard Generalized Markup Language encoding standard (SGML) designed specifically for marking up information contained in archival finding aids. Finding aids are documents that describe the content of collections. For example, archival fond, print and photo libraries and manuscript collections available in libraries and museums. Finding aids usually have two components, an *Intellectual* component that describes the interrelationships between a group of records and the administrative entities that created them and a *physical* component that is used to find the actual physical items in the collection (Theodoridou and Doerr, 2001).

5.1.1 Introduction to EAD

The EAD structure information consists of two segments:

1. One segment that provides information about the finding aid itself (refers to the title, compiler, and compilation date use to find the related information). It then splits this segment in 2 sub-elements, they are:
   - EAD Header <eadheader>
   - Front Matter <frontmatter>
2. The second segment <archdesc> provides specific information about a given archival object held by or group of object. (a collection, a record group, or a series), it consists of the actual description of the archival materials and associated administrative and contextual information.

In the following sections, the description for the EAD elements is based on the Encoded Archival Description Tag Library (EAD – Index by Element Name, 2002).

5.1.1.1 Structure of <eadheader> in EAD:

<eadheader>: This is used for retrieval and its contents provide a brief citation to the finding aid (the title, compiler, and compilation date use to find the related information). It is also used by repository staff. Elements required in the <eadheader> are in the following order: <eadid>, <filedesc>, <profiledesc> (optional), and <revisiondesc> (optional). These elements and their sub-elements provide useful documenatory information.

- <eadid> (ead id): A unique code designated for a particular EAD finding aid document.
- <filedesc> (file description): Covers much of the bibliographic information about the finding aid, including its author, title, subtitle, and sponsor (all in the <titlestmt>), as well as the edition, publisher, publishing series, and related notes. Required elements are: <titlestmt>, <editionstmt> (optional), <publicationstmt> (optional), <seriesstmt> (optional), <notestmt> (optional).
- <profiledesc> (profile description): Covers information about the creation of the encoded version of the finding aid, including the name of the agent, place, and date of encoding. The <profiledesc> element also designates the predominant and minor languages used in the finding aid. The required sub-elements include: <langusage> and <creation>.

- <revisiondesc> (revision description): Describes information about changes or alterations that have been made to the encoded finding aid. The revisions may be recorded as part of a <list> or as a series of <change> elements.

Figure 16 describes the hierarchy structure of <eadheader> using SGML.

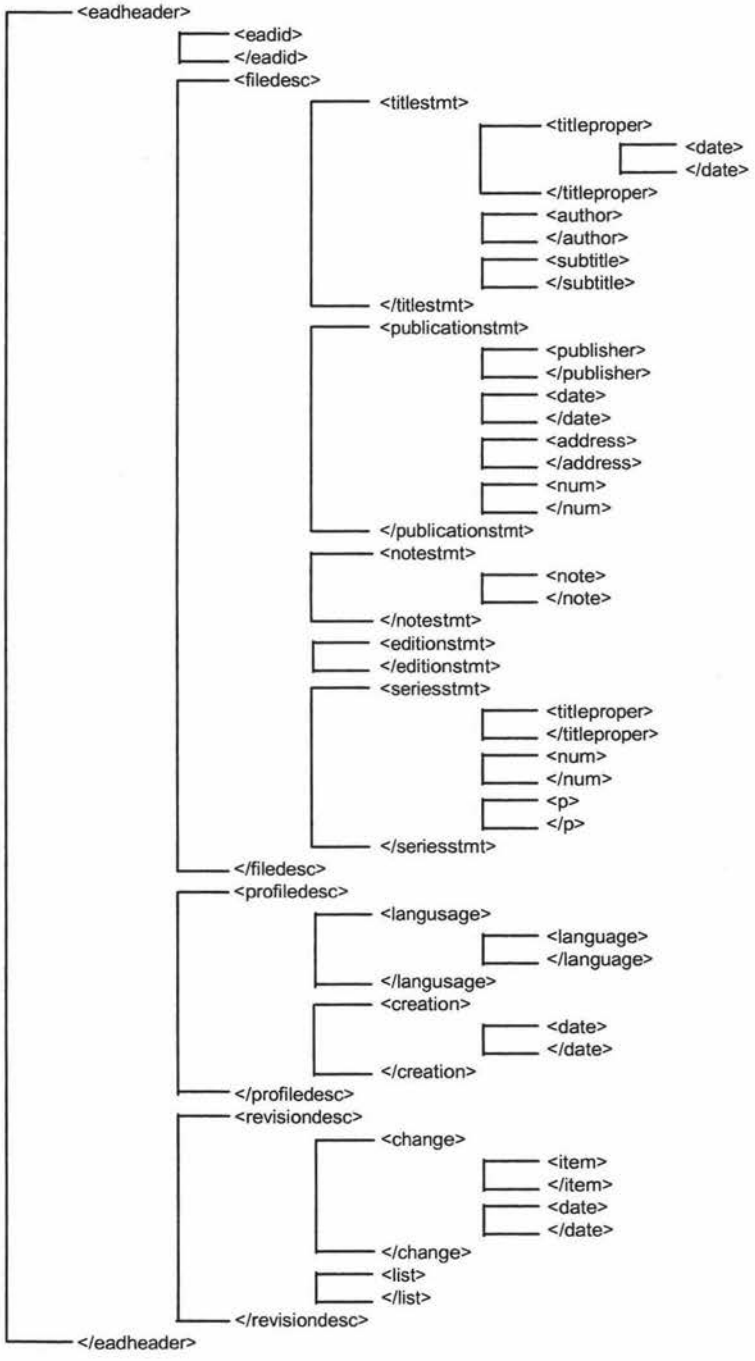


Figure 16 Hierarchy Structure of <eadheader>

5.1.1.2 Structure of <frontmatter> in EAD:

<frontmatter>: It focuses on the creation, publication, or use of the finding aid. Examples include: title page, preface, dedication, and instructions for using the finding aid. <frontmatter> is being converted as the type of "String" in EAD when mapping <frontmatter> to ERD, it can be described as:

$$< frontmatter >= E_{31}Document \ : \ (p_3 has\ note) \ : \ E_{62}String^{[: p_2 \ has\ type : E_{55}Type \ (frontmatter)]}.$$

5.1.1.3 Structure of < archdesc > in EAD:

<archdesc>: Describes the content, context, and extent of a body of archival material, including administrative and supplementary information. To provide a more detailed view of the consistent parts; information is organized in unfolding hierarchical levels. The required sub-elements for <archdesc> are: <did>, <scopecontent> and <dsc>. In order to provide an initial basic description of the material the <did> element appear in <archdesc> before more detailed descriptions are presented in <scopecontent> and <dsc>.

- <did> (descriptive identification): It identifies core information about the described material. The sub-elements of <did> may contain: <abstract>, <head>, <note>, <origination>, <physdesc>, <physloc>, <repository>, <unitdate>, <unitid>, <unittitle>.
- <controlaccess> (controlled access headings): This is a wrapper element containing key access points for the described material and enables searching across a computer network. It records information of individual or institutional ownership of the described material, such as the name and address.
- <dscgrp> (description of subordinate components): A wrapper element that covers information about the hierarchical groupings of the described material. The subcomponents can be presented in several different forms or levels of descriptive detail. These subcomponents can be mapped as entity "Man-Made Object", such as: $[c, c01....c0n] = E_{22}Man\text{-}Made\ Object$ and:

  $$[c, c01....c0n] = E_{22}Man\text{-}Made\ Object \ : \ (P_{46}is\ composed\ of) \ : \ E_{22}Man\text{-}Made\ Object .$$

Figure 17 shows the hierarchy structure of <archedesc> using SGML:

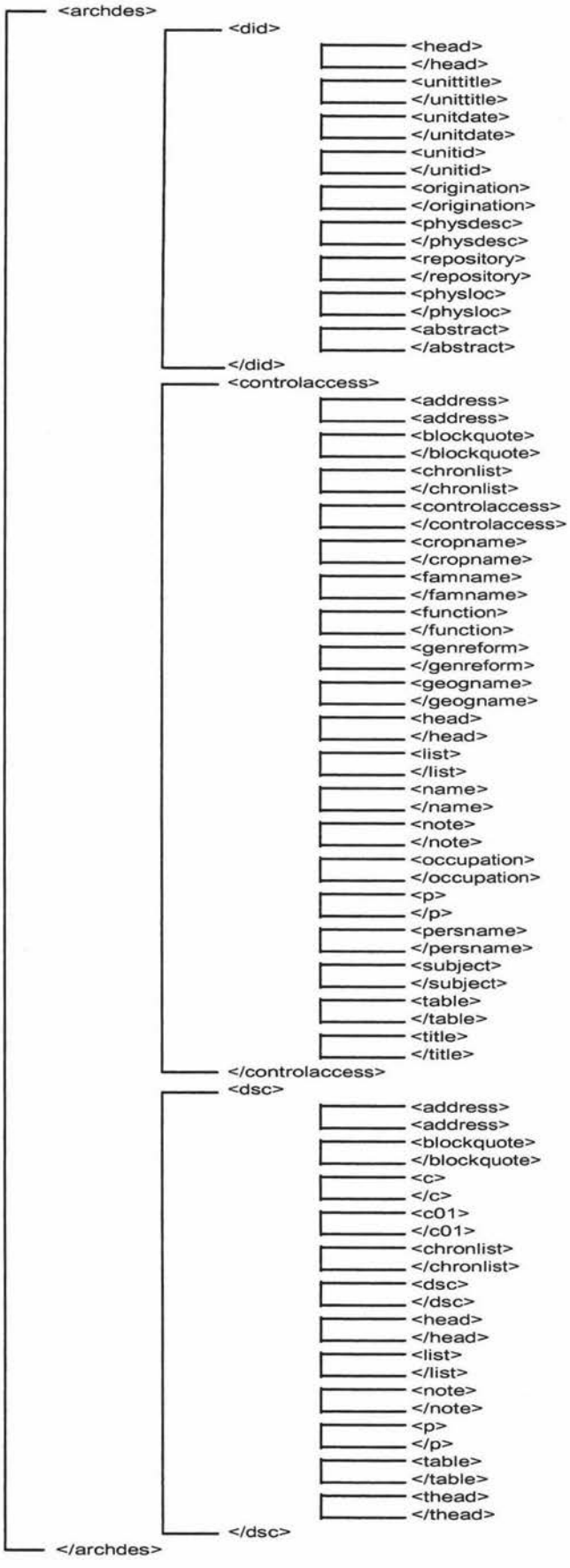


Figure 17 Hierarchy Structure of <archedesc>

5.1.2 Mapping Formalism

The EAD is an encoding finding aid which provides fully described information of primary source material. CRM is a domain ontology in the form of object-oriented semantic model. The aim of CRM is to solve the problem of semantic interoperability between various kinds of museum data.

Before examining the mapping between the EAD DTD and the CRM, the major features of EAD DTD and CRM are discussed.

5.1.2.1 Differences between EAD DTD and CRM

As Theodoridou and Doerr (2001) points out, the purpose of the EAD DTD is to define the structure of the EAD document. There are few required elements contained in EAD DTD, the rest of the elements are optional, thus, having a fully described document depends on intellectual and financial considerations.

These are three types of elements in EAD (Theodoridou and Doerr, 2001):

- **Wrapper element** does not contain text directly; the text is stored in the sub element/nested element instead.
- **Generic element** refers to the elements that can be located in more than one place with the EAD. For example, 'date' is regarded as the generic element, which may be found in several elements such as 'publication' element or the 'creation' element.
- **Formatting element** refers to the element used to invoke a special character or formatting of the text, such as line spaces, emphasis and underline.

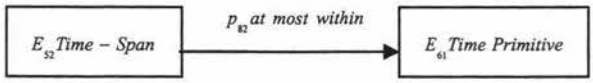
The CRM is a semantic model based on a formal ontology. CRM is different from other types of data structure; because as an ontology it is able to reflect and preserve the semantic meaning of the original documentation. However, in EAD DTD, instances of the archival data need to be fitted into the required element structure. The structure is designed for data capture and retrieval using the minimum of documentation to optimise coding. As Doerr (2001b) states, applying a relative flat structure database does not allow rich and meaningful information to be stored. In CRM, the entity identifier is required to be independent from the content of the data, for example, the identifier of the object may be separated from the rest of the object

content, but in EAD DTD, instances of the data identifier and their content are set into the elements within the whole record. Theodoridou and Doerr (2001, p3) points out that, *"the aim in the mapping of EAD DTD to CRM is not the description of the DTD structure in terms of the CRM but rather to create a semantic equivalent of the information about the world described in a EAD DTD instance."*

5.1.2.2 Mapping scheme

CRM properties (links) are represented by the unique id of the applicable entity (domain), the name of the link and the name of the referred entity (range).

Consider the following diagram:

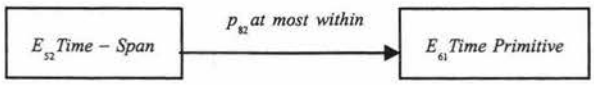


Notation below expresses this relationship:

$\langle p_{82} \rangle : E_{52} \rightarrow E_{61}$

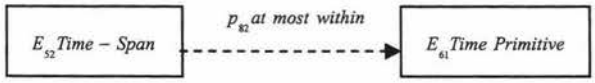
Applying the notation proposed in the previous section, the above formula can be written in the following formats according to the cardinality relationship between the entity and the link:

Consider the following diagrams:



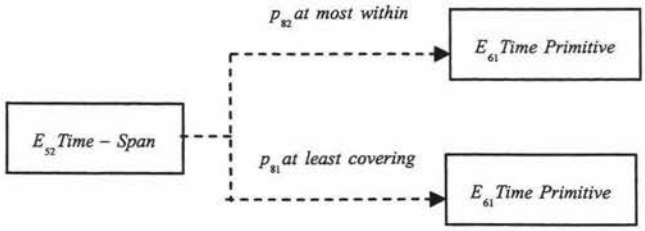
Notation:

$\langle p_{82} \rangle : E_{52} \rightarrow E_{61}$



Notation:

$[\![p_{82}]\!] : E_{52} \rightarrow \Box E_{61}$



Notation:
$[p_{82}, p_{81}]: E_{52} \rightarrow E_{61} \oplus E_{61}$

If the link is inherited from other entities, and no links can be found between the entities (refer to 4.2.4 for more details), the expression could be written:

$\langle eadheader \rangle = E_{31} Document : \overline{p_{106} is\ composed\ of} : E_{31} Document$.

For those "property on property" relationships, the formula is:

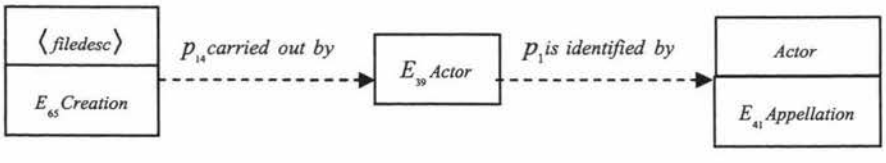
$E_1 CRM\ Entity : p_3 has\ note^{[p_{111} has\ type: E_{55} Type]} : E_{62} String$, and formula can be expressed as:

$\langle p_3, p_{111} \rangle : E_1 \rightarrow E_{62} \times E_{55}$, $[p_3, \langle p_3, p_{111} \rangle] : E_1 \rightarrow E_{62} \oplus (E_{62} \times E_{55})$ or

$[\![\langle p_3, p_{111} \rangle]\!] : E_1 \rightarrow \square (E_{62} \times E_{55})$.

After choosing the mapping format, each EAD elements can be mapped to the corresponding entity in CRM:

- The content of the EAD element can be used to create a correspondent unique identifier in CRM, such as: $\langle Address \rangle = E_{45} Address$; and from this mapping, additional mapping can be derived from it:
  $\langle Address \rangle = E_{45} Address(has\ note : String)$.
- Sometimes, the content of the element may also used to derive additional CRM entities. For example, $\langle titleproper \rangle\ \langle date \rangle\ \langle / date \rangle\ \langle / titleproper \rangle$ can be expressed as the mapping of $\langle titleproper \rangle$ and $\langle date \rangle$.
- For the nested element mapping to CRM, there is: $\langle ead \rangle = E_{31} Document$, for it's nested element $\langle eadheader \rangle$, there is:
  $\langle eadheader \rangle = E_{31} Document : \overline{p_{106} is\ composed\ of} : E_{31} Document$.
- If the relation expressed by a EAD element relates to a path with intermediate entities in the CRM, it is a "join" relationship, this is illustraded in the following diagram:

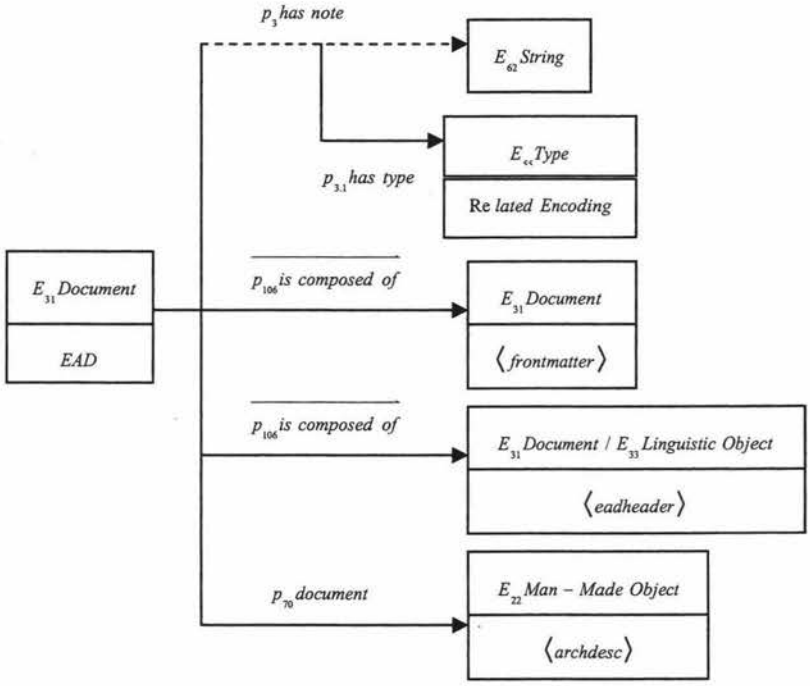


In this case, $E_{39} Actor$ is the intermediate entity.

5.1.3 Mapping of the Encoded Archival Description (<ead>) tag

The record for a finding aid instance can be mapped to the CRM Document: $\langle ead \rangle = E_{31} Document$. According to the structure of EAD, the $\langle ead \rangle$ itself contains three sub-elements: $\langle eadheader \rangle$, $\langle frontmatter \rangle$ and $\langle archdesc \rangle$, which are all required elements in $\langle ead \rangle$.

Mapping diagram:



Notation:

$$\left\langle [\![\langle p_3, p_{3.1} \rangle]\!], \left(\overrightarrow{p_{106}}\right)^2, p_{106}, p_{70} \right\rangle : E_{31} \rightarrow \square(E_{62} \times E_{55}) \times (E_{31})^2 \times E_{33} \times E_{22}.$$

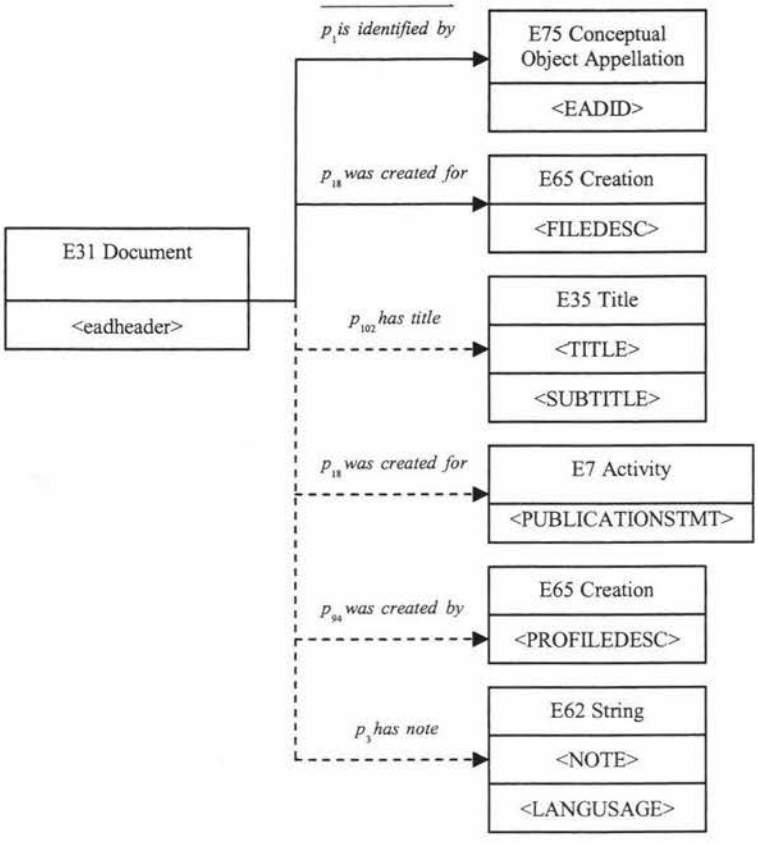
(Note: in this notation, the expression of using the power of the entity: $\left(\overrightarrow{p_{106}}\right)^2 : E_{31} \rightarrow (E_{31})^2$ is that these two pairs of mappings share the same link and the two EAD elements map to the same entities as well. This will apply to the rest of the mapping in the following sections.)

5.1.4 Mapping of the elements of <eadheader>

5.1.4.1 Mapping of <eadheader>:

The mapping of <eadheader> can be divided in two parts; the first is mapping the <eadheader> to $E_{31} Document$, the second is mapping the <eadheader> to $E_{33} Linguistic\ Object$. See the following two diagrams:

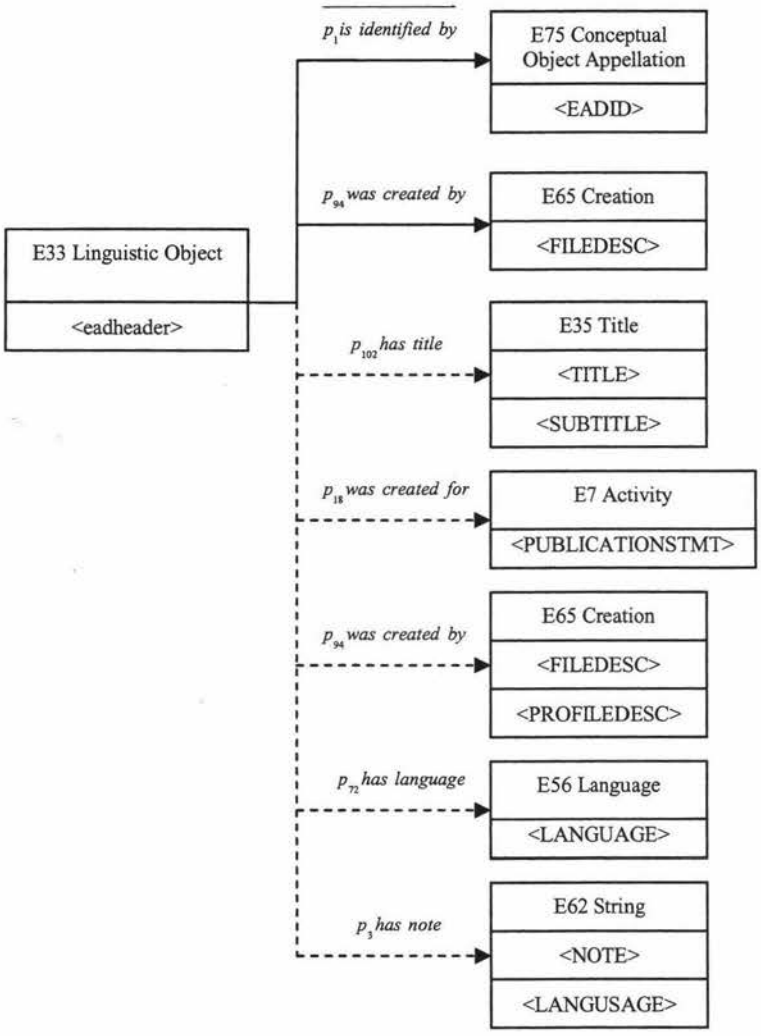
Mapping diagram: when $\langle eadheader \rangle = E_{31} Document$.



Notation:

$$\left\langle \overline{p_1}, p_{18}, [\![p_{102}, p_{18}, p_{94}, p_3]\!] \right\rangle : E_{31} \rightarrow E_{75} \times E_{65} \times \square\left(E_{35}, E_7, E_{65}, E_{62}\right).$$

Mapping diagram: when $\langle eadheader \rangle = E_{33} Linguistic\ Object$.



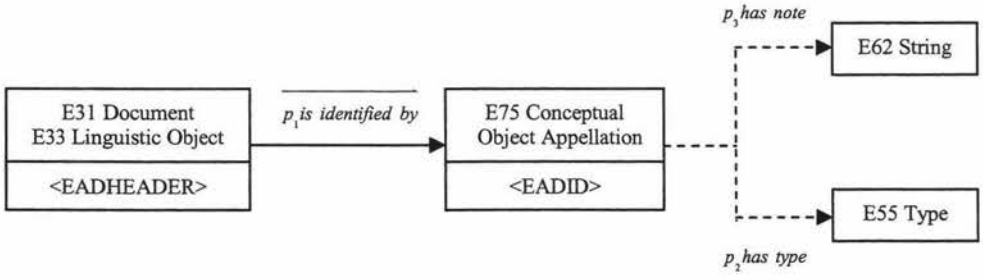
Notation:

$$\left\langle \overline{p_1}, p_{18}, [\![p_{102}, p_{18}, p_{94}, p_{72}, p_3]\!] \right\rangle : E_{31} \rightarrow E_{75} \times E_{65} \times \square(E_{35}, E_7, E_{65}, E_{56}, E_{62}).$$

5.1.4.2 Mapping of <eadid>:

< *eadid* > is a required element in < *ead* > that designates a unique code for a particular EAD finding aid document. The relationship between <eadid> and <eadheader> can be mapped as:

Mapping diagram:



Notation:

$\langle p_1 [\![p_3, p_2]\!] \rangle : E_{31} \rightarrow E_{75} \times \square (E_{62}, E_{65})$

5.1.4.3 Mapping of <titleproper>, <subtitle>:

<titleproper> is a required element within the <titlestmt> (subelement of <filedesc>). Part of <eadheader>, <subtitle> is a secondary or subsidiary name of the encoded finding aid and is subordinate to the main name encoded in <titleproper>. Both of <titleproper> and <subtitle> can be mapped to $E_{35} Title$ in CRM, different "has type" link has been applied to them in order to distinguish between the two titles:

$< eadheader > . < filedesc > . < titlestmt > . < titleproper > \ = E_{31} \ Document \ :$

$P_{102} \ has \ title^{[: P_2 \ has \ type: E_{55} \ Type \ (proper)]} : E_{35} \ Title$

$< eadheader > . < filedesc > . < ftitlestmt > . < titleproper > \ = E_{31} \ Document \ :$

$P_{102} \ has \ title^{[: P_2 \ has \ type: E_{55} \ Type \ (subtitle)]} : E_{35} \ Title$

Notation

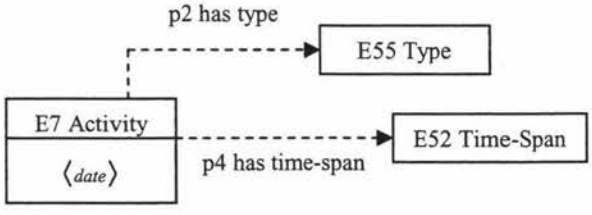
$[\![\langle p_{102}, p_{102.1} \rangle]\!] : E_{31} \rightarrow \square (E_{55} \times E_{35})$.

5.1.4.4 Mapping of <date>:

In EAD, <date> is a generic element that contains a month, day, or year in any format. Examples of dates that might merit encoding are a person's birth date, the date the materials were acquired, or the date of an event in a chronology. These dates may be entered in the form of text or numbers, and may consist of a single date or range of dates.

<date> maps to $E_7 Activity$ in CRM. The following diagram describes the mapping of <date> in CRM, because <date> is regarded as a generic element, the mapping of the <date> is independent of other EAD elements.

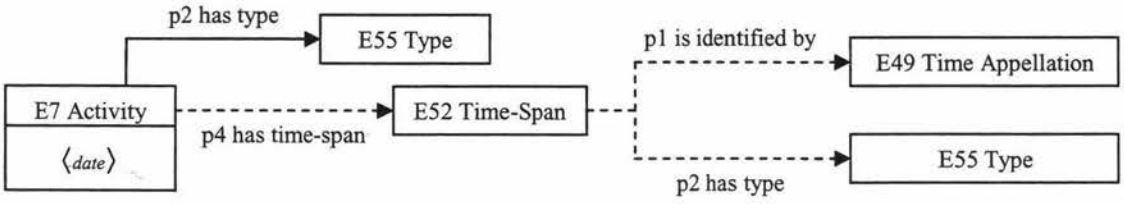
Mapping diagram:



According to CRM, $E_{52} Time-Span$ is able to link with the entities of:

' E_{49}*Time Appellation* ', ' E_{41}*Appellation* ', ' E_{55}*Type* ', ' E_{62}*String* ', ' E_{61}*Time Primitive* ', ' E_{54}*Dimension* ', and ' E_{52}*Time−Span* ' in an optional way. Here is an example of using ' E_{49}*Time Appellation* ' and ' E_{55}*Type* ' to represent <date>:

Mapping diagram:

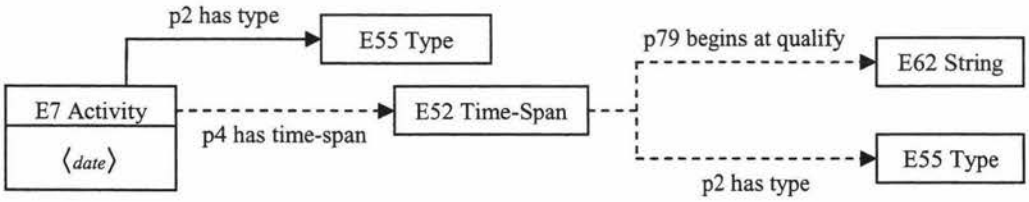


Notation:

$$\langle p_2, [\![p_4 [\![p_1, p_2]\!]]\!] \rangle : E_7 \rightarrow E_{55} \times (\square(E_{52} \times \square(E_{49}, E_{55})))$$

Furthermore, we can also use the other properties to describe the entity E_{52}*Time − Span*, such as:

Mapping diagram:

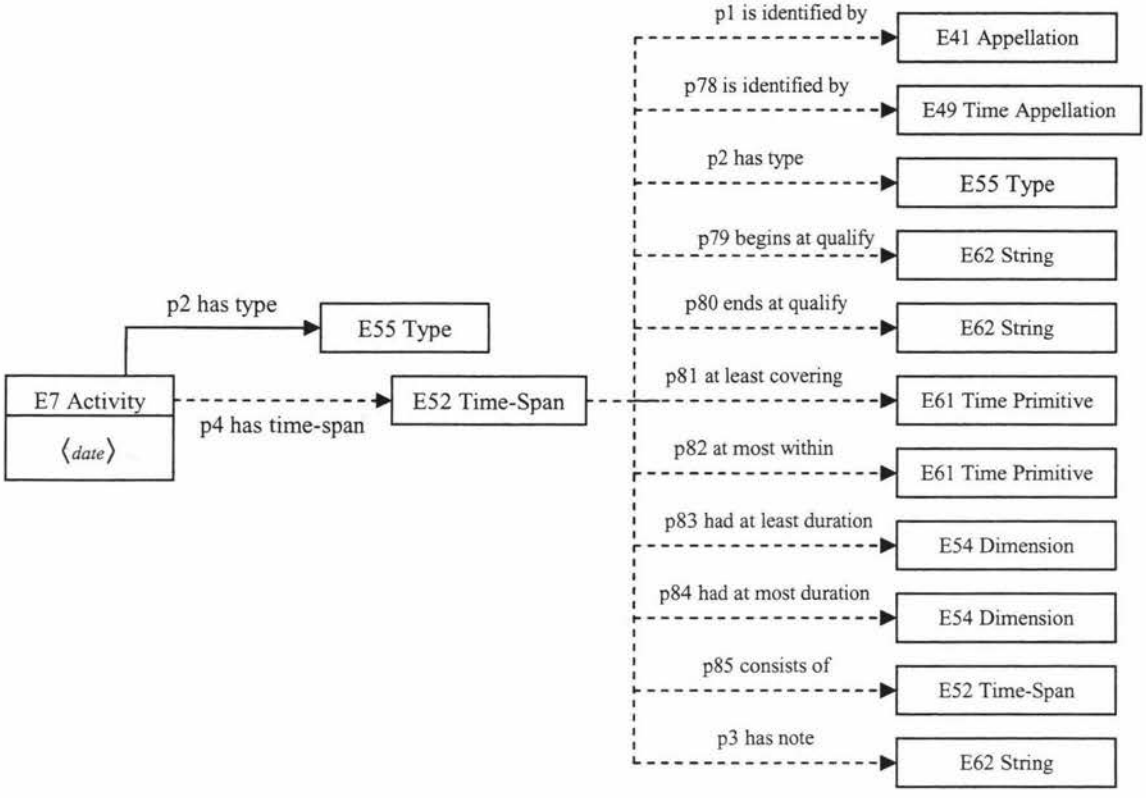


Notation:

$$\langle p_2, [\![p_4 [\![p_{79}, p_2]\!]]\!] \rangle : E_7 \rightarrow E_{55} \times (\square(E_{52} \times \square(E_{62}, E_{55})))$$

Following is a summary of the possible descriptions of "date" in terms of using E_7*Activity* :

Mapping diagram:



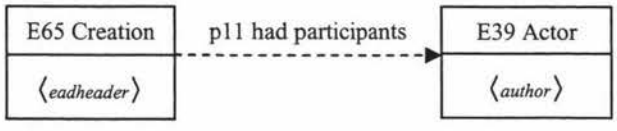
Notation:

$$[\![p_2, [\![p_4, \langle p_4, [\![p_1, p_{78}, p_2, p_{79}, p_{80}, p_{81}, p_{82}, p_{83}, p_{84}, p_{85}, p_3]\!]\rangle]\!]\!] : E_7$$
$$\rightarrow \Box(E_{55}, \Box(E_{52}, E_{52} \times \Box(E_{41}, E_{49}, E_{55}, E_{62}, E_{62}, E_{61}, E_{61}, E_{54}, E_{54}, E_{52}, E_{62})))$$

5.1.4.5 *Mapping of <author>:*

The <author> element is available in the <titlestmt> portion of the <eadheader>. It represents name(s) of institution(s) or individual(s) responsible for compiling the intellectual content of the finding aid. It may include a brief statement indicating the nature of the responsibility, for example, archivist, collections processor, or records manager. <author> maps to $E_{39}Actor$ in CRM:

Mapping Diagram:



Notation:

$$[\![p_{11}]\!] : E_{65} \rightarrow \Box E_{39}$$

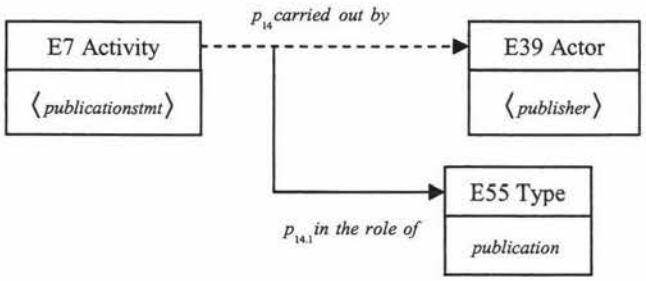
5.1.4.6 Mapping of <publisher>:

Two situations apply to <publisher>:

1. When used in the <publicationstmt> portion of <eadheader> and in the <titlepage> element in <frontmatter>, the <publisher> is the name of the party responsible for issuing or distributing the encoded finding aid.
2. When used in the <imprint> section of a Bibliographic Reference <bibref>, the <publisher> is the name of the party issuing a monograph or other bibliographic work cited in the finding aid.

"publisher" maps to $E_{39}Actor$ in CRM, because there's no specific activity for publication in CRM, so the entity $E_7Activity$ can be used to represent 'publication' here.

Mapping diagram:



Notation:

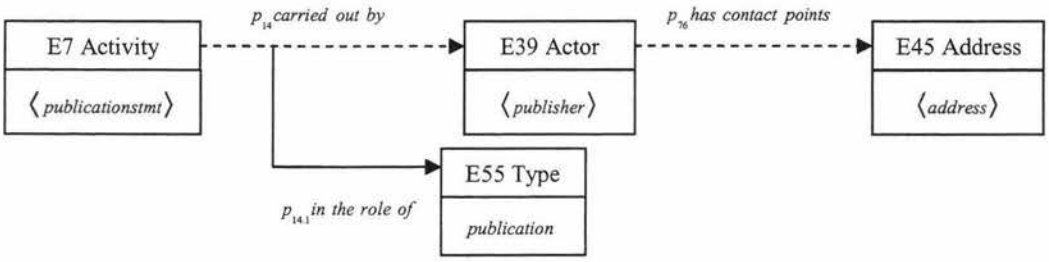
$[\![\langle p_{39}, p_{55}\rangle]\!] : E_7 \rightarrow \square(E_{39} \times E_{55})$

5.1.4.7 Mapping of <address>:

<address> is a generic element for information about the place where someone or something is located and may be reached. Examples include a postal address for a repository, or the electronic mail address and phone number of the party granting publication permission. <address> maps to $E_{45}Address$ in CRM

As <address> is a generic element, it may occurred within elements such as <publicationstmt>, <titlepage> and <note>. In the following mapping, <address> is treated as a sub-element of <publicationstmt>:

Mapping diagram:



This leads to the following expression: $[\![\langle p_{14}, p_{14.1}\rangle p_{76}]\!] : E_7 \rightarrow \square((E_{39} \times E_{55}) \times E_{41})$.

5.1.4.8 Mapping of <publicationstmt>:

The <publicationstmt> is a wrapper element within the <filedesc> portion of <eadheader>, it may contains just text, laid out in paragraphs in which case it is mapped to a plain string, or it may include the sub elements of publisher, address, date and number, <publicationstmt> can be mapped to $E_7 Activity$:

$< ead >.< eadheader >.< filedesc >.< publicationstmt > = E_{31}\ Document:$
$p_{15}\ was\ taken \text{int} o\ accout\ by\ :\ E_7\ Activity$

Notation:

$[\![p_{15}]\!] : E_{31} \rightarrow \square E_7$.

5.1.4.9 Mapping of <publisher>, <date>, <address> and <num>:

As mentioned above, the <publicationstmt> may include the sub elements of <publisher>, <address>, <date> and <num>, which allow for more specific tagging of a publisher's name and address, the date, and the number of publication. These elements map CRM as:

- <publicationstmt>.<publisher> :

  $< publicationstmt >.\ < publisher >= E_7\ Activity\ :\ P_{14}\ carried\ out\ by: E_{39} Actor\ .$

  Notation:

  $[\![p_{14}]\!] : E_7 \rightarrow \square E_{39}$

- <publicationstmt>.<date>:

  $< publicationstmt >.\ < date > = E_7 Activity: p_4\ has\ time\text{-}span: E_{52}\ Time\ Span\ .$

  Notation:

  $[\![p_4]\!] : E_7 \rightarrow \square E_{52}$

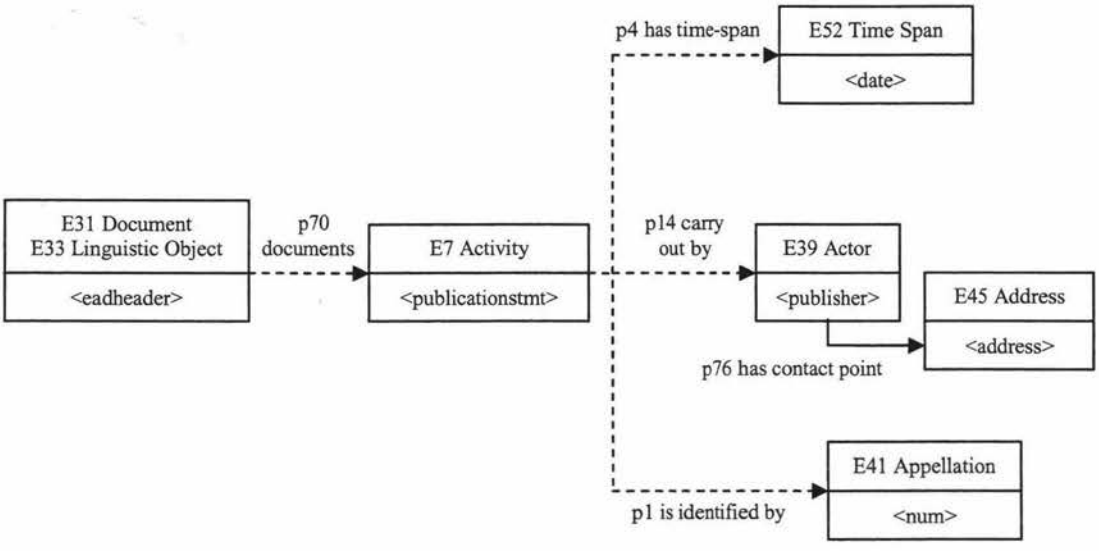
- <publicationstmt>.<address>: refer to the mapping of <address> stated in the previous section 5.1.4.7
- <publicationstmt>.<num>:

  $< publicationstmt > . < num > = E_7 Activity : p_1 \text{ is identified by } : E_{41} \text{ Appellation}$

  Notation:

  $[\![p_1]\!] : E_7 \rightarrow \square E_{41}$

Mapping diagram shows the relationship between <eadheader> and <publicationstmt>.



Notation:

$[\![p_{70}, [\![p_4, [\![\langle p_{14}, p_{76} \rangle]\!], p_1]\!]]\!] : E_{31} / E_{33} \rightarrow \square(E_7 \times \square(E_{52}, (E_{39} \times E_{49}), E_{41}))$

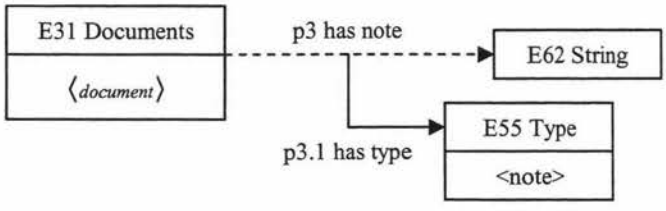
5.1.4.10 Mapping of <note> and <editionstmt>:

- <note>

  <note> is the sub-element of <notestmt>, it can be mapped to $E_{55} Type$.

  Mapping of <note> is:

  Mapping diagram:



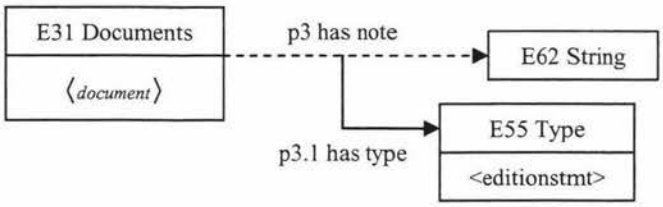
  Notation:

  $[\![\langle p_3, p_2 \rangle]\!] : E_{31} \rightarrow \square(E_{62} \times E_{55})$.

- <editionstmt>

<editionstmt> is an optional elements within the <filedesc> portion of the <eadheader> element that groups information about a finding aid edition. The information they contain is useful as display information only, and so it is mapped to the "has note" attribute of the <ead>.<eadheader> document:

Mapping diagram:



Notation:

$[\![\langle p_3, p_2 \rangle]\!] : E_{31} \rightarrow \square(E_{62} \times E_{55})$.

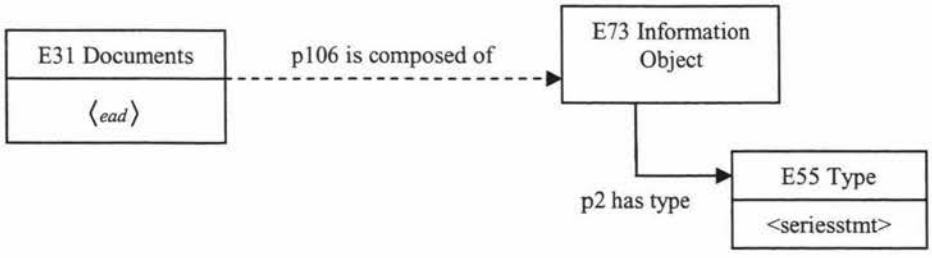
5.1.4.11 *Mapping of <seriesstmt>, <titleproper>, <num> and <p>:*

<seriesstmt> is a wrapper element within the <filedesc> portion of <eadheader> that groups information about the published monographic series. The <seriesstmt> may contain just text, laid out in paragraphs <p>, or it may include the sub-elements of <titleproper>, <num> and <p>, which allow for more specific tagging of names or numbers associated with the series. Following are the mappings of these elements:

- <seriesstmt>

  <seriesstmt> maps to $E_{55}Type$ in CRM, mapping for <seriesstmt> can be expressed as:

  Mapping diagram:



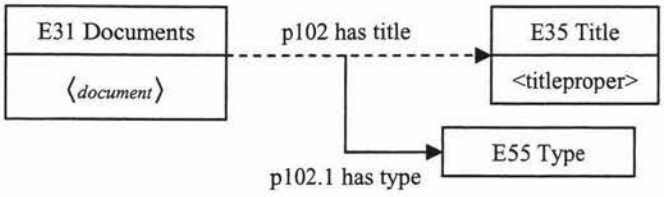
  Notation:

  $[\![p_{106}\langle p_2 \rangle]\!] : E_{31} \rightarrow \square(E_{73} \times E_{55})$.

- <titleproper>

  <titleproper> maps to $E_{35}Title$, mapping for <titleproper> can be expressed as:

Mapping diagram:



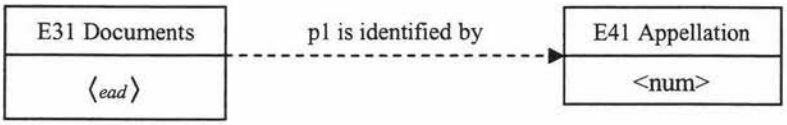
Notation:

$[\![\langle p_{102}, p_{102.1}\rangle]\!] : E_{31} \rightarrow \square(E_{35} \times E_{55})$.

- <num>

<num> maps to E_{41} *Appellation*, mapping of <num> is:

Mapping diagram:



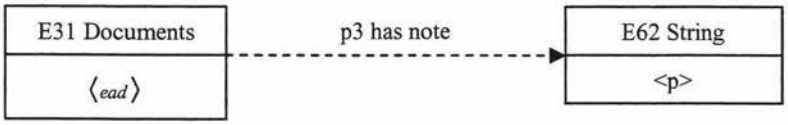
Notation:

$[\![p_1]\!] : E_{31} \rightarrow \square E_{41}$.

- <p>

<p> maps to E_{62} *String*, mapping for <p> is:

Mapping diagram:



Notation:

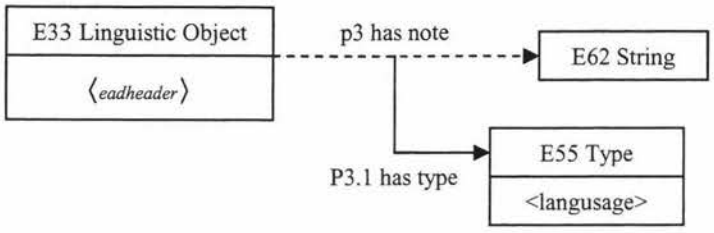
$[\![p_3]\!] : E_{31} \rightarrow \square E_{62}$.

5.1.4.12 Mapping of <langusage>, <language>:

- <langusage>

<langusage> is an optional sub-element within the <profiledesc> portion of the <eadheader> that specifies the language or communication system in which the finding aid is written. <langusage> maps to E_{55} *Type*:

Mapping diagram:

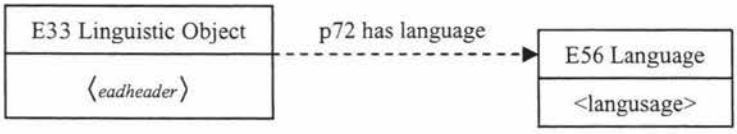


Notation:

$[\![\langle p_3, p_{3.1} \rangle]\!] : E_{33} \rightarrow \square(E_{55} \times E_{62})$.

- <language>

<language> is an optional sub-element within the <profiledesc> portion of the <eadheader> that provides a statement about languages, sublanguages, and dialects represented in an encoded finding aid. The language(s) in which the finding aid is written can be further specified using the <language> sub-element within <langusage>. <language> maps to CRM as "Language", and mapping of the <language> and <language> in EAD are as follow:



Notation:

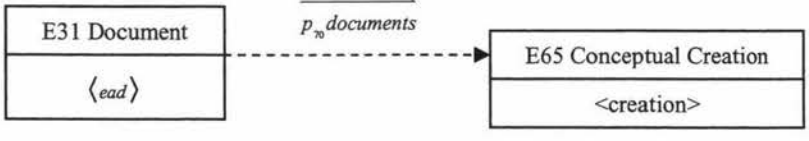
$[\![p_{72}]\!] : E_{33} \rightarrow \square E_{56}$.

5.1.4.13 Mapping of <creation>:

- <creation>

<creation> is a sub-element of the <profiledesc> part of <eadheader>, it includes the information about the person(s) or agency(ies) responsible for the encoding, the date, and the circumstances.

Mapping diagram:



Notation:

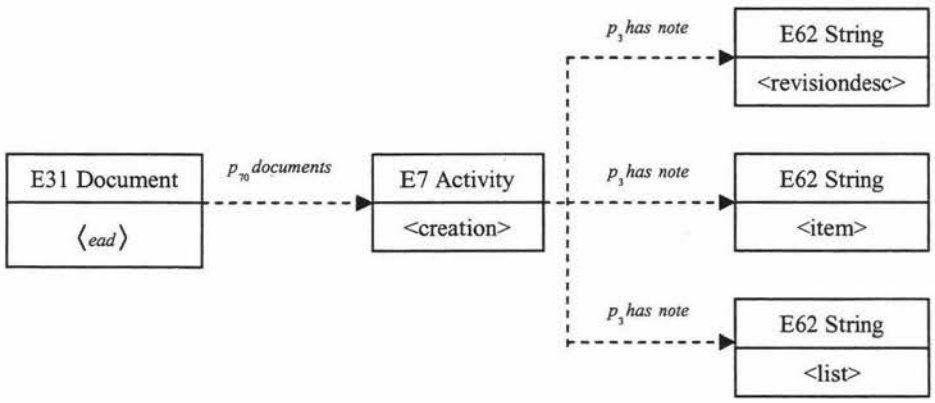
$[\![\overline{p_{70}}]\!] : E_{31} \rightarrow \square E_{65}$.

5.1.4.14 Mapping of <revisiondesc>, <item>, and <list>:

<revisiondesc> is an optional sub-element of the <eadheader> for information about changes or alterations.

<item>, <date> and <list> are all the sub-elements of <revisiondes> (the sub-element of <change>). In <revisiondesc> element, the <item> describes information about a revision to the finding aid and the <item> can be a number, word, or phrase. <list> is used as a formatting element that contains a series of words or numbers separated from one another and arranged in a linear, often in a vertical sequence. Here are the mappings of <revisiondesc>, <item>, and <list>:

Mapping diagram:



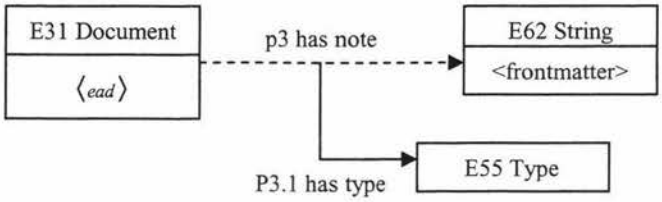
Notation:

$[\![p_{70}, \langle p_{70}, p_3 \rangle]\!] : E_{31} \rightarrow \square(E_7, E_7 \times E_{62})$

5.1.5 Mapping of the elements of <frontmatter>:

The <frontmatter> focuses on the creation, publication, or use of the finding aid rather than information about the materials being described. Examples include title page, preface, dedication, and instructions for using the finding aid. <frontmatter> maps as $E_{62}String$ of <ead> ($E_{31}Document$), all the contents of <frontmatter> contain in the $E_{62}String$, so there's no detailed mapping for each of the <frontmatter> element.

Mapping diagram:

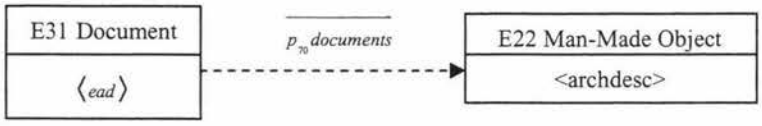


Notation:

$[\![\langle p_3, p_{3.1}\rangle]\!] : E_{31} \rightarrow \square(E_{55} \times E_{62})$

5.1.6 Mapping of the elements of <archdesc>:

<archdesc> is a wrapper element for the bulk of an EAD document instance, which describes the content, context, and extent of a body of archival materials, including administrative and supplemental information that facilitates use of the materials. <archdesc> maps to $E_{22} Man-Made\ Object$. The mapping of it is:



Notation:

$[\![\overline{p_{70}}]\!] : E_{31} \rightarrow \square(E_{22})$

5.1.6.1 Mapping of <did>.<head> and <did>.<unittitle>:

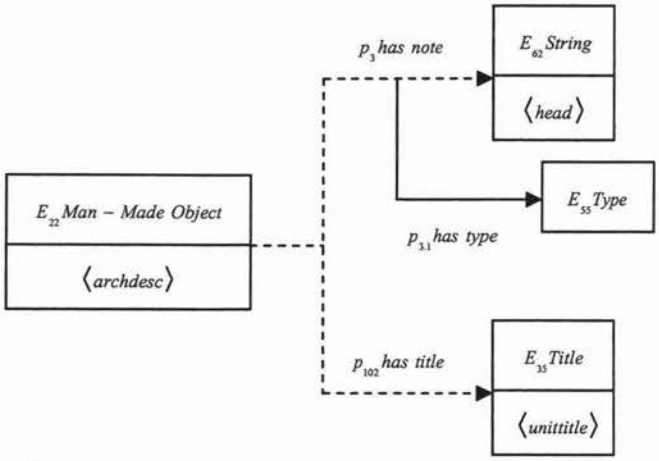
- <head>

  <head> is a generic element that designates the title or caption for a section of text, including a list. It can be mapped to $E_{62} String$.

- <unittitle>

  As an important sub-element of <did>, the <unittitle> is used to encode the name of the described materials. <unittitle> can be mapped to $E_{35} Title$.

Mapping diagram for <head> and <unittitle>:



Notation:

$$[\![[\![\langle p_3, p_{3.1}\rangle]\!], p_{102}]\!] : E_{22} \rightarrow \square\big((E_{62} \times E_{55}), E_{35}\big)$$

5.1.6.2 *Mapping of <did>.<unitdate>:*

As mentioned above, <unitdate> describes the creation date of the described material. In order to map <unitdate>, the 'Production' with a Time-Span is introduced in the mapping:

Mapping diagram:

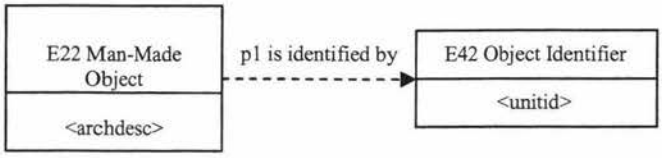
Notation:

$$[\![p_{31}[\![p_4]\!]]\!] : E_{22} \rightarrow \square\big(E_{12} \times (\square E_{52})\big)$$

5.1.6.3 *Mapping of <did>.<unitid>:*

<unitid> refers to any alpha-numeric text string that serves as a unique reference point or control number for the described material, it maps to E_{42}*Object Identifier* in EAD. The following diagram and notation shows the relationship with <archdesc>:

Mapping diagram:



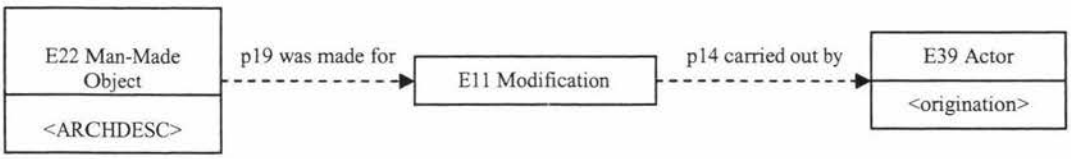
Notation:

$[\![p_1]\!] : E_{22} \rightarrow \square E_{42}$

5.1.6.4 Mapping of <did>.<origination>:

<origination> refers to information about the individual or organization responsible for the creation, accumulation or assembly of the described materials before their incorporation into an archival repository. <origination> maps to E_{39} *Actor* in EAD. The entity E_{11} *Modification* is created as an intermediate in this mapping. The following diagram shows an example of the relationship between <archdesc> and <origination>:

Mapping diagram:



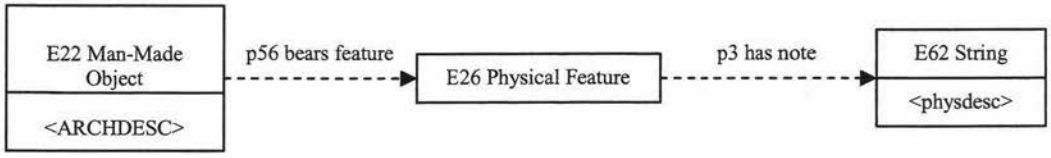
Nation:

$[\![p_{19}, \langle p_{19}, p_{14} \rangle]\!] : E_{22} \rightarrow \square (E_{11}, E_{11} \times E_{39})$

5.1.6.5 Mapping of <did>.<physdesc>:

<physdesc> is a wrapper element for bundling information about the appearance or construction of the described materials. Some of the aspects of the appearance can be mapped to CRM entity directly. However, some of the aspects do not have a straight forward mapping, such as the aspects of colour, style, etc. To map these aspects to CRM, Theodoridou and Doerr (2001) choose to map all this information to E_{62} *String* in CRM, and the entity E_{26} *Physical Feature* is created as an intermediate of this mapping.

Mapping diagram:



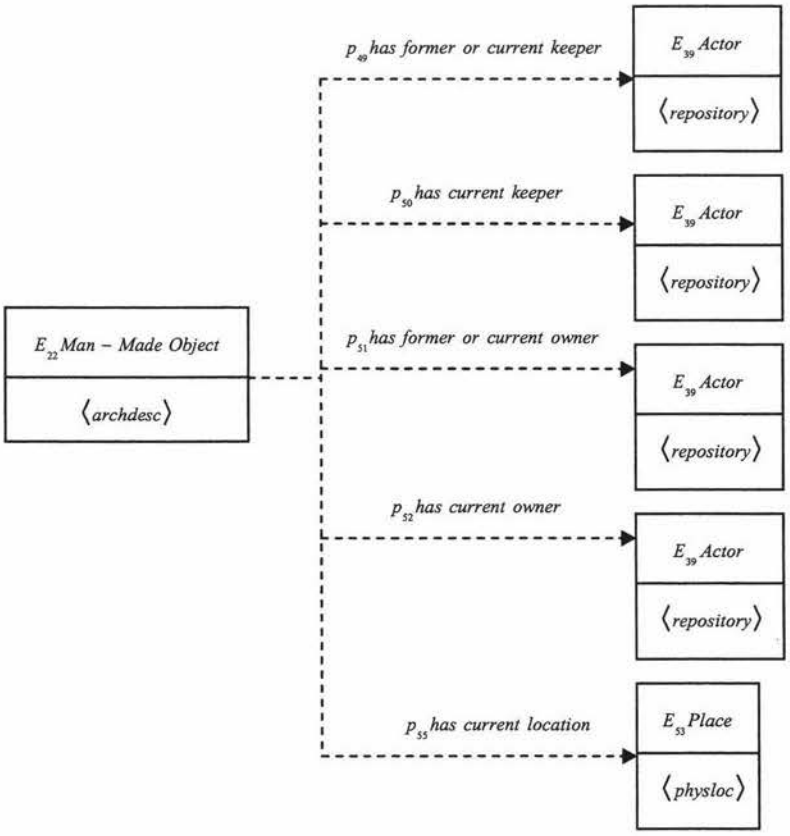
Notation:

$[\![p_{56}, \langle p_{56}, p_3 \rangle]\!] : E_{22} \rightarrow \square(E_{26}, E_{26} \times E_{62})$

5.1.6.6 Mapping of <did>.<repository> and <did>.<physloc>:

<repository> covers the information of individual or institution that provides intellectual accessing to the described material. <physloc> provides information for identifying the place where the described materials are stored. In the mapping to CRM, <repository> maps to entity E_{39} *Actor*, <physloc> maps to entity E_{53} *Place*, mapping of <repository> and <physloc> are described in the following diagram:

Mapping diagram:



Notation:

$[\![p_{49}, p_{50}, p_{51}, p_{52}, p_{55},]\!] : E_{22} \rightarrow \square(E_{39}, E_{39}, E_{39}, E_{39}, E_{55})$

5.1.6.7 Mapping of <did>.<abstract> and <did>.<note>:

- <abstract>

<abstract> is a brief summary of the materials being described, used primarily to encode bits of biographical or historical information about the creator and statements, such as the scope, content, arrangement.

- <note>

<note> is a generic element that provides a short statement explaining the text, indicating the basis for an assertion, or citing the source of a quotation or other information.

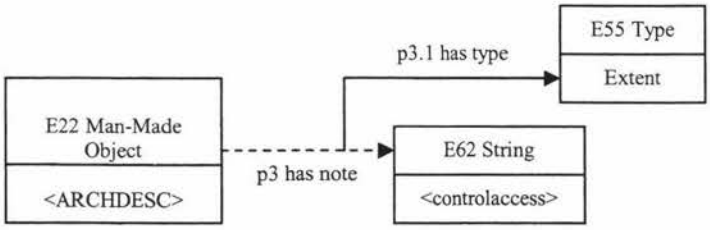
Both <abstract> and <note> map to $E_{62}String$ in CRM, thus they share the same Notation:

$$[\![p_3]\!]: E_{22} \rightarrow \square(E_{62})$$

5.1.6.8 Mapping of <controlaccess> elements

<controlaccess> is a wrapper element that designates key access points for the described materials and enables authority-controlled searching across finding aids on a computer network, such as the name, address of the individual/institute, etc. Usually, this element map to entity $E_{62}String$ in EAD:

Mapping diagram:



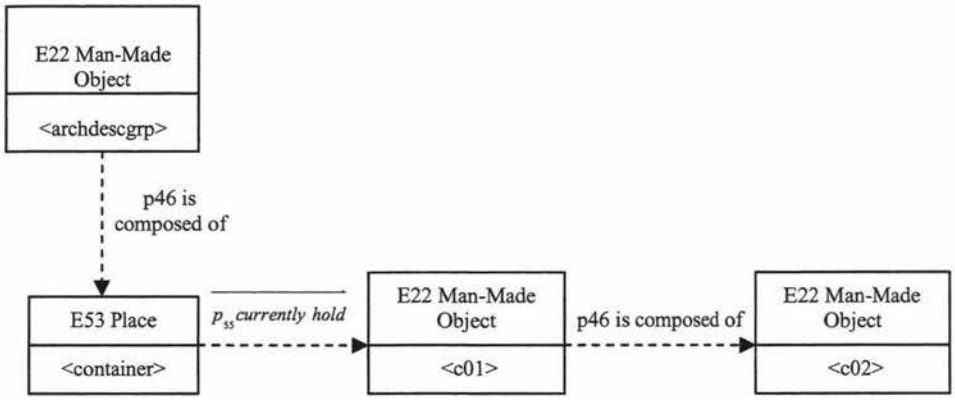
Notation:

$$[\![\langle p_3, p_{3.1}\rangle]\!]: E_{22} \rightarrow \square(E_{55} \times E_{62})$$

5.1.6.9 Mapping of <dscgrp> elements

<dscgrp> is a wrapper element used within the <archdescgrp> sub-element of <eadgrp> in the EAD. The <dscgrp> can be both numbered (01-12) and unnumbered provide information about the content, context and extent of a subordinate body of materials. The following diagram shows the mapping of <dscgrp> elements, the numbers of the <dscgrp> can be extended from c0n to c12n.

Mapping diagram:



Notation:

$$[\![p_{46}[\![\overline{p_{55}}[\![p_{46}]\!]]\!]]\!]: E_{22} \rightarrow \square\left(E_{53} \times \square(E_{22} \square (E_{22} \times \square E_{22}))\right).$$

5.2 Mapping of the Dublin Core Element set to the CIDOC CRM (based on Dublin Core Version 1.1 and CRM Version 3.0)

This is the second of the three international validation exercises examined in this thesis. The source of this validation exercise is "Mapping the Dublin Core Metadata Element Set to the CIDOC CRM" (Doerr, 2000).

5.2.1 Introduction to Dublin Core

The Dublin Core metadata element set is a standard for cross-domain information resource description. The basic set consists of 14 elements, these are displayed in the following table (Table 2), published by the DC Metadata Initiative (Dublin Core – Metadate Element Set, Version 1.1, 2003). If used to map information from a typical record of a cultural object in a museum / art gallery collections system considerable amount of information would be lost. However, the purpose of this validation exercise is to determine whether the required DC data can in fact be mapped to the CRM. The mapping exercise is expressed using both the graphical and proposed category theory notation.

Table 2 The DC Element (DC Metadata Initiative (2003)

Element Name	Label	Definition	Comment:
Title	Title	A name given to the resource.	Typically, Title will be a name by which the resource is formally known
Creator	Creator	An entity primarily responsible for making the content of the resource.	Examples of Creator include a person, an organization, or a service. Typically, the name of a Creator should be used to indicate the entity.
Subject	Subject and Keywords	A topic of the content of the resource.	Typically, Subject will be expressed as keywords, key phrases or classification codes that describe a topic of the resource. Recommended best practice is to select a value from a controlled vocabulary or formal classification scheme.
Description	Description	An account of the content of the resource.	Examples of Description include, but is not limited to: an abstract, table of contents, reference to a graphical representation of content or a free-text account of the content.
Publisher	Publisher	An entity responsible for making the resource available	Examples of Publisher include a person, an organization, or a service. Typically, the name of a Publisher should be used to indicate the entity
Contributor	Contributor	An entity responsible for making contributions to the content of the resource.	Examples of Contributor include a person, an organization, or a service. Typically, the name of a Contributor should be used to indicate the entity.
Date	Date	A date of an event in the lifecycle of the resource.	Typically, Date will be associated with the creation or availability of the resource. Recommended best practice for encoding the date value is defined in a profile of ISO 8601 [W3CDTF] and includes (among others) dates of the form YYYY-MM-DD.
Type	Resource	The nature or genre of the content of the resource.	Type includes terms describing general categories, functions, genres, or aggregation levels for content. Recommended best practice is to select a value from a controlled vocabulary (for example, the DCMI Type Vocabulary [DCT1]). To describe the physical or digital manifestation of the resource, use the FORMAT element.
Format	Format	The physical or digital manifestation of the resource	Typically, Format may include the media-type or dimensions of the resource. Format may be used to identify the software, hardware, or other equipment needed to

			display or operate the resource. Examples of dimensions include size and duration. Recommended best practice is to select a value from a controlled vocabulary (for example, the list of Internet Media Types [MIME] defining computer media formats).
Identifier	Resource Identifier	An unambiguous reference to the resource within a given context	Recommended best practice is to identify the resource by means of a string or number conforming to a formal identification system. Formal identification systems include but are not limited to the Uniform Resource Identifier (URI) (including the Uniform Resource Locator (URL)), the Digital Object Identifier (DOI) and the International Standard Book Number (ISBN).
Source	Source	A Reference to a resource from which the present resource is derived.	The present resource may be derived from the Source resource in whole or in part. Recommended best practice is to identify the referenced resource by means of a string or number conforming to a formal identification system.
Language	Language	A language of the intellectual content of the resource	Recommended best practice is to use RFC 3066 [RFC3066] which, in conjunction with ISO639 [ISO639]), defines two- and three-letter primary language tags with optional subtags. Examples include "en" or "eng" for English, "akk" for Akkadian", and "en-GB" for English used in the United Kingdom
Relation	Relation	A reference to a related resource.	Recommended best practice is to identify the referenced resource by means of a string or number conforming to a formal identification system.
Coverage	Coverage	The extent or scope of the content of the resource.	Typically, Coverage will include spatial location (a place name or geographic coordinates), temporal period (a period label, date, or date range) or jurisdiction (such as a named administrative entity). Recommended best practice is to select a value from a controlled vocabulary (for example, the Thesaurus of Geographic Names [TGN]) and to use, where appropriate, named places or time periods in preference to numeric identifiers such as sets of coordinates or date ranges
Rights	Rights Management	Information about rights held in and over the resource.	Typically, Rights will contain a rights management statement for the resource, or reference a service providing such information. Rights information often encompasses Intellectual Property Rights (IPR), Copyright, and various Property Rights. If the Rights element is absent, no assumptions may be made about any rights held in or over the resource.

5.2.2 Mapping Formalism

5.2.2.1 Comparison between DC and CRM

The Dublin Core, an information resource based on a set of core data extracted from the original sources using a finding aid. Doerr (2000) states that the value of the Dublin Core is in its simplicity and its interoperability in assigning all kinds of data resources. Generally speaking, there are no essential restrictions to the types of resources to which Dublin Core metadata can be applied (Dublin Core – Metadate Element Set, Introduction Section, 2003). Dublin Core is trying to extend the semantic meaning of the original source defintions. However, as Doerr (2000) points out that the extendibility of Dublin Core appears to be based on the concept, "the looser the definition, the more flexible the interpretation and the wider its application" (Doerr, 2000, p3). Such as extending the meaning of "object" to cover both "man-made object" and "conceptual object", the extendibilty will not make sense when applying the fields of "man-made object" such as "has dimension", "has current location" to "conceptual object".

The CRM, as discussed before, is a semantic model based on a formal ontology. Doerr (2000) claims that the CRM provides the generosity and interoperability by constraining extensibility, however, the Dublin Core acquires its generosity by applying "underspecified" notions. On the other hand CRM provides "constraint extension" to its structure, by adding a well defined abstraction to the original source; this is achieved using the following mechanisms (Doerr, 2000):

- The new sub-classes must be created only if its parents-class entity existing in CRM.
- The new sub-properties must be created only if its parents-properties existing in CRM.
- As an intermediate entity in the relationship, it also regarded as a subclass of the existing entities in this relationship.

For the mapping of Dublin Core to CRM, Doerr (2000) summarizes three advantages, they are:

1. The mapping is able to combine the greater granularity of Dublin Core and get more restricted definitions from the CRM, provided that the related Dublin Core resource falls within the scope of CRM.

2. The Dublin Core structure is a standard that is regarded by Doerr (2001b) as being too small to fulfil advanced requirements and the relative flat structure may cover lots of hidden constraints. By mapping the DC to CRM, enables the recovery of the rich information that has been "flattened out" in the Dublin Core record.
3. CRM is very flexible relationship involving two entities; this allows information to be transferred between two sides of the entities. Such as, a movie with Tom Hanks in Dublin Core could be also represented as an attribute of Tom Hanks.

5.2.2.2 *Mapping scheme*

Similar to the mapping scheme in EAD mapping, CRM links are represented by the unique id of the applicable entity, the name of the link and the name of the referred entity, the formula below is used to express this relationship:

$E_{52}Time-Span : p_{82}at\ most\ within : E_{61}Time\ Primitive$

Besides using the above method of representation, the researcher also applies the notation proposed in the EAD. By applying the category theory notation, the above formula can be written in one of several ways.

According to the cardinality relationship between the entity and the link:

$\langle p_{82} \rangle : E_{52} \rightarrow E_{61}$, $[\![p_{82}]\!] : E_{52} \rightarrow \Box E_{61}$ or $[p_{82}] : E_{52} \rightarrow E_{61}$

However, if the link is inherited from other entities, and no links can be found between the entities (refer to 4.2.4 for more details), the expression becomes:

$\overline{[\![p_{106}]\!]} : E_{19} \rightarrow \Box E_{19}$

For those links of links, the formula is:

$E_1 CRM\ Entity : p_3 has\ note^{[p_{111} has\ type: E_{55} Type]} : E_{62} String$, based on the cardinality relationship between entities. The formula can be expressed as: $\langle p_3, p_{111} \rangle : E_1 \rightarrow E_{62} \times E_{55}$, $[p_3, \langle p_3, p_{111} \rangle] : E_1 \rightarrow E_{62} \oplus (E_{62} \times E_{55})$

or $[\![\langle p_3, p_{111} \rangle]\!] : E_1 \rightarrow \Box (E_{62} \times E_{55})$.

After choosing the mapping format, the Dublin Core elements can be mapped to the corresponding entity in CRM. Here is an example of a Dublin Core structure record for 'painting' from Doerr (2000):

```
painting{
        integer record _ id;
        string title;
        string kind _ of _ title;
        structure pa int er{
                        stirng artist _ name;
                        stirng nationality;
                        string contribution;
                        } pa int er
        time creation _ date _ begin;
        time creation _ date _ end;
        string creation _ date _ comment;
        string creation _ place;
        string last _ exhibition;
                        }painting;
```

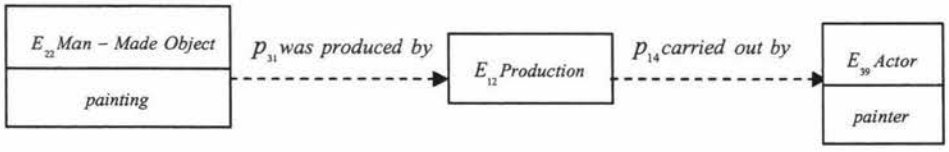
In this example, the whole record of the 'painting' is treated as an instance of $E_{22}Man-Made\ Object$, therefore, the DC fields of the 'painting' can be mapped to a correspondence entity value in CRM, such as $last\_exhibition$. This corresponds to the entity $E_7 Activity$ in CRM. The $E_7 Activity$ can also be used to derive additional attributes, such as $E_7 Activity$ is carried out by $E_{39}Actor$.

Furthermore, to express the relationship between $last\_exhibition$ and 'painting', the following description can be applied:

$painting.last\_exhibition = E_{22}Man-Made\ Object : p_{16}was\ used\ for : E_7 Activity$.

In this case, the property 'was used for' is used to build the connection between $last\_exhibition$ ($E_7 Activity$) and 'painting' ($E_{22}Man-Made\ Object$).

If the relation expressed by a DC field relates to a path with intermediate entities in the CRM, it is a "join" relationship, the following diagram illustrates this relationship:



In this case, $E_{12}Production$ is the intermediate entity.

5.2.3 Mapping Dublin Core Metadata Element Set to CRM

The mapping of the Dublin Core in this thesis is based on Dublin Core Metadate Element Set Version 1.1 (2003), as showed in Table 2, the Dublin Core metadata elements are:

- Title
- Creator
- Subject
- Description
- Publisher
- Contributor
- Date
- Type
- Format
- Identifier
- Source
- Language
- Relation
- Right

The following mapping sections are based on this sequence of the metadata elements and the equivalent definition of the above elements are quoted from the Dublin Core Metadata Element Set, Version 1.1 (2003).

5.2.3.1 *Mapping of the Dublin Core described resource.*

The DCMI Type Vocabulary (2003) is also regarded as [DCT1], it represents the nature or genre of the content of the Dublin Core elements, it covers description of general categories, functions, genres, or aggregation levels for content. Generally speaking, the source of the Dublin Core elements is based on the DCMI Type Vocabulary. These controlled terms are: collection, dataset, event, image, interactive resource, physical object, service, software, sound and text. Doerr (2000) has mapped the following DCMI Type terms to corresponding CRM entities, this mapping is shown in the following table (Table 3):

Table 3 DCMI Type Vocabulary ([DCT1]) and their corresponding CRM entities

DCMI Type Terms	CRM Entities
collection	E_{22} *Man – Made Object*
dataset	E_{73} *Information Object*
event	E_{7} *Activity*
image	E_{38} *Image* or E_{23} *Iconographic Object*
interactive resource	E_{73} *Information Object*
model	E_{29} *Design or Procedure*
party	E_{39} *Actor*
physical object	E_{19} *Physical Object*
place	E_{22} *Site*
service	E_{39} *Actor*
software	E_{73} *Information Object*
sound	E_{73} *Information Object*
text	E_{33} *Linguistic Object*

(Note: according to Doerr (2000), the DCMI Type "service" usually is controlled and followed under the instruction of human beings, so service maps to Actor in all the common cases. However, if it is necessary to differentiate between people and the service programme, the mapping needs to be extended.)

In the following mapping, when the value of the Dublin Core element is restricted by DCMI Type Vocabulary, Doerr (2000) suggests to express this as DC [DCT1, DCMI Type Vocabuary], such as DC[DCT1 event].

5.2.3.2 *Dublin Core Qualifiers*

When the restricted resources of the Dublin Core elements are associated with Dublin Core Qualifier, (Dublin Core Qualifier is terms extended from the fifteen DC elements that serve to inform user and programs how to interpret the value in the DC element properly). Doerr (2000) suggests expressing it as: DC.DC Elements.Qualifier,

such as the qualifier 'Alternative' is the qualifier for element 'Title', so here is the expression: "DC.Title.Alternative". In the following sections, descriptions of the DC Qualifiers are based on the Dublin Core Qualifiers (2000).

5.2.3.3 Mapping of DC.Title

DC.Title refers to the name given to the resource.

- Based on Doerr (2000), an 'Appellation' can be treated as 'Title' when there are no constraints to DC.Title, in this case, he maps DC.Title as 'Appellation'. This leads to the following mapping and notation:

  $DC.Title = E_{41} Appellation$

  $DC > DC.Title = E_1 CRM\ Entity : p_1 is\ identified\ by : E_{41} Appellation$

  Notation:

  $[\![p_1]\!] : E_1 \rightarrow \square E_{41}$.

  When DC.Title has a specific meaning, Doerr (2000) interprets DC.Title as 'Title' in the specific meaning for human creation, and interprets DC.Title as 'Title' in all the other cases, such as 'Appellation'.

- In the first case, DC.Title is restricted by the source of human creation, it generates the following mapping:

  $Titlte = E_{35} Title$

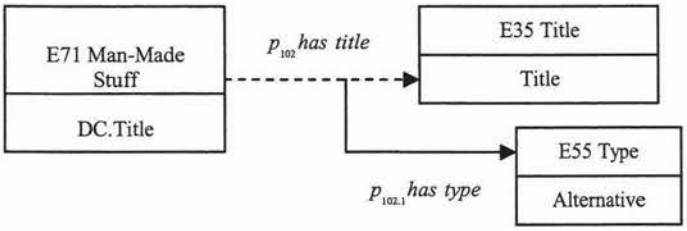
  $DC > DC.Title = E_{71} Man-Made\ Stuff : p_{102} has\ title : E_{35} Title$

  Notation:

  $[\![p_{71}]\!] : E_{102} \rightarrow \square E_{35}$.

- In the second situation when DC.Title is restricted by DC Qulifier 'Alternative':

  Mapping diagram:



  Notation:

  $[\![p_{102}, p_2]\!] : E_{71} \rightarrow \square (E_{35} \times E_{55})$.

5.2.3.4 Mapping of DC.Agents

The concept of Agent in Dublin Core is regarded as Actor in CRM, there are three Dublin Core elements which can be regarded as Agent, they are: DC.Creator, DC.Contributor and DC.Publisher.

Qualifiers for CCP (Creator / Contributor / Publisher) are: Agent Type, Agent Name, Agent Affiliation, Agent role and Agent Identifier. (Iannella, 1999).

The following mappings are applied:

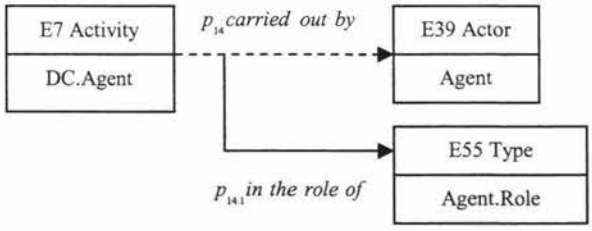
- *Agent Type* = E_{55}*Type*
- *Agent Name* = E_{41}*Appellation*

In this case, the E_{41}*Appellation* used to map both 'Agent Name' and 'DC.Title'. This is because the repetition of identical properties is encouraged in CRM structure, and creates added meaning (Doerr, 2000).

- *Agent Affiliation* = E_{40}*Legal Body*
- *Agent Role* = E_{55}*Type*

Event, action and activity are elements in CRM associated with 'Agent Role'. When mapping the 'Agent Role' to CRM, Agent maps to E_7*Activity* :

Mapping diagram:



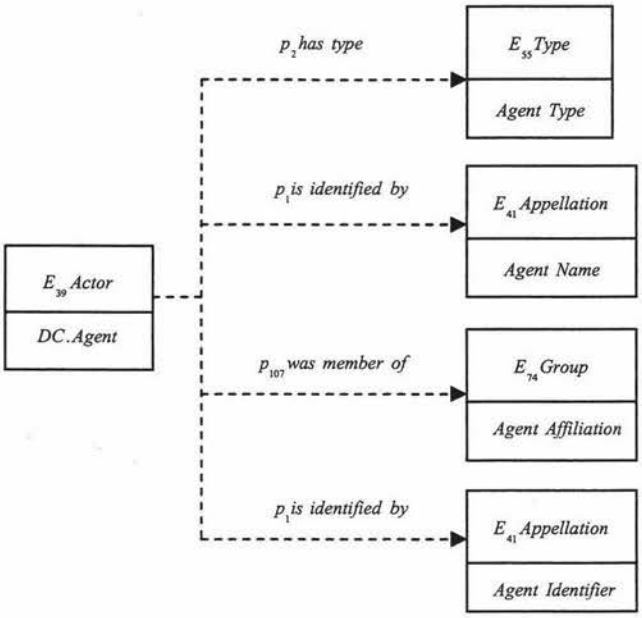
Notation:

$[\![p_{14}, p_{14.1}]\!] : E_7 \rightarrow \square(E_{39} \times E_{55})$

- *Agent Identifier* = E_{41}*Appellation*

For the mapping of Agent Type, Agent Name, Agent Affiliation and Agent Identifier, Agent maps to E_{39}*Actor* :

Mapping diagram:



Notation:

$[\![p_2, p_1, p_{107}, p_1]\!]: E_{39} \rightarrow \square(E_{55}, E_{41}, E_{74}, E_{41})$

(Note: In this diagram, $E_{74}Group$ is the superclass of $E_{40}Legal\ Body$)

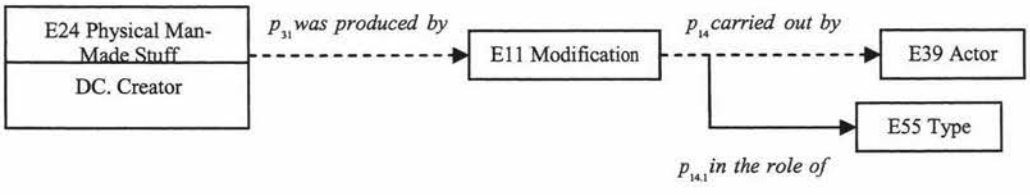
5.2.3.5 Mapping of DC.Creator and DC.Contributor

DC.Creator refers to an entity primarily responsible for the content of the resource. DC.Contributor refers to an entity responsible for making contributions to the content of the resource. The mapping of Creator and Contributor are very similar, the property "in the role of" can be used to make the distinction between them. However, in terms of creators, there is still the need to distinguish the difference between physical stuff and conceptual stuff. The mapping of DC.Creator is categorized into three cases:

- When the DC.Creator is associated with $E_{24}Physical\ Man-Made\ Stuff$:

  The entity of 'Modification' has been used in the mapping. Modification has been regarded as the most common property for physical stuff, and the relevant pre-existing parts are used and rearranged.

  Mapping diagram:



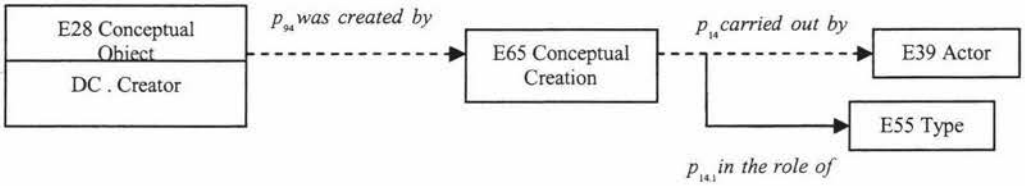
Notation:

$$[\![p_{31},[\![\langle p_{14},p_{14.1}\rangle]\!]]\!]: E_{24} \rightarrow \square(E_{11} \times \square(E_{39} \times E_{55})).$$

- The DC.Creator is associated with E_{28}*Conceptual Object*

  Compared with physical stuff, there is no trace of a pre-existing part for conceptual object, so no Modification is required in the mapping of Creator to Conceptual Object.

  Mapping diagram:

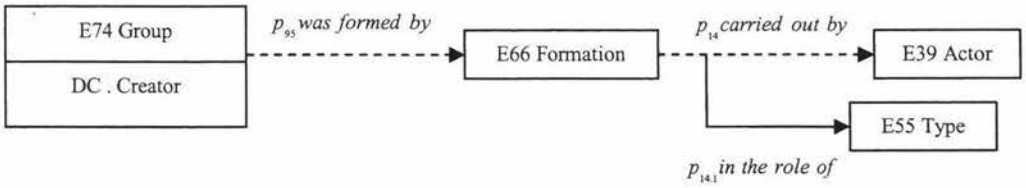


Notation:

$$.[\![p_{94},[\![\langle p_{14},p_{14.1}\rangle]\!]]\!]: E_{28} \rightarrow \square(E_{65} \times \square(E_{39} \times E_{55}))$$

- The DC.Creator is associated with E_{74}*Group*

  Physical Stuff and Conceptual Object may not cover the entire situation in the DC creation world. Groups are formed to imply formation. For example, the formation of a government. A physical person is not seen as the product of the creation process. Following diagram shows the mapping:



Notation:

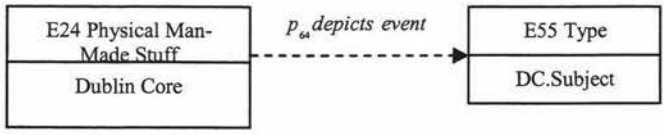
$$[\![p_{95},[\![\langle p_{14},p_{14.1}\rangle]\!]]\!]: E_{74} \rightarrow \square(E_{66} \times \square(E_{39} \times E_{55}))$$

5.2.3.6 Mapping of DC.Subject

DC.Subject refers to a topic of the content of the resource. Doerr (2000) believes that only 'Physical Man-Made Stuff' and 'Conceptual Object' are the two resources that have a subject in the proper sense, so when Dublin Core is related to E_{24}*Physical Man – Made Stuff* and E_{28}*Conceptual Object* , DC.Subject is associated with one of them:

- When DC is related to E_{24}*Physical Man – Made Stuff* :

Mapping diagram:

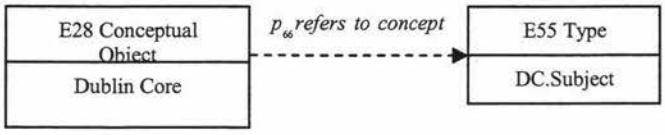


Notation:

$[\![p_{64}]\!]: E_{24} \rightarrow \square E_{55}$.

- When DC is related to E_{28} *Conceptual Object*:

  Mapping diagram



  Notation:

  $[\![p_{66}]\!]: E_{28} \rightarrow \square E_{55}$.

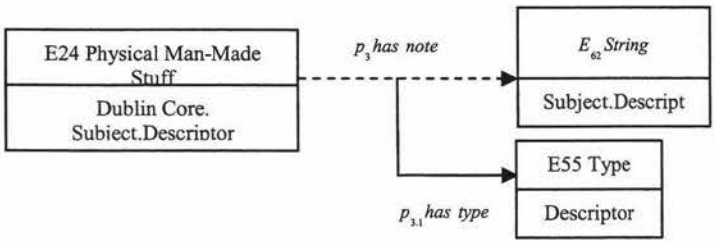
The qualifiers elements of DC.Subject are:

- Subject Descriptor
- Subject Classification
- Subject Keywords

Following is the mappings for the subject qualifier elements:

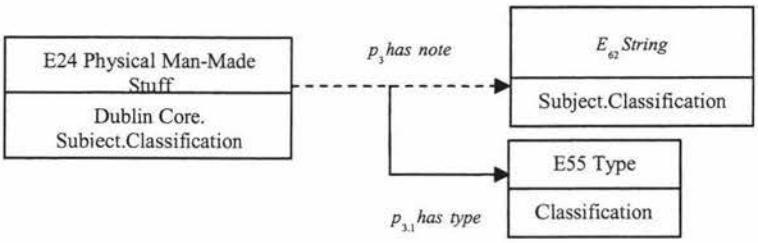
- *Subject Descriptor* = E_{55}*Type*

  The following diagram shows the mapping of Subject Descriptor:

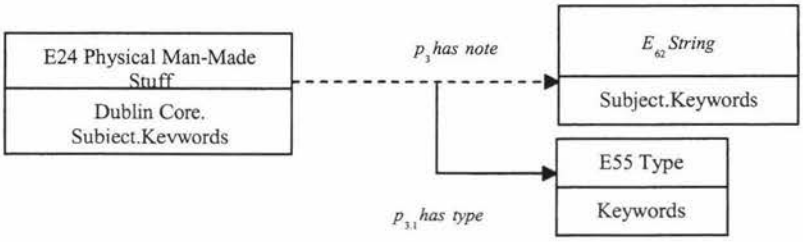


- *Subject Classification* = E_{55}*Type*

  The following diagram shows the mapping of Subject.Classification:



- *Subject Keywords* = E_{55}*Type*



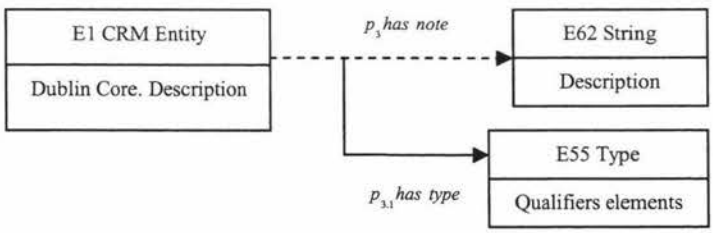
Notation (all have the same):

$$[\![\langle p_3, p_{3.1} \rangle]\!] : E_{24} \rightarrow \square(E_{62} \times E_{55})$$

5.2.3.7 Mapping of DC.Description

DC.Description refers to an account of the content of the resource. Same with Dublin Core, every entity in CRM has a description, so DC.Description can be mapped as E_{62}*String*, and DC.Description is associated with E_1*CRM Entity*.

The following diagram shows the mapping of DC.Description:



Notation for the above diagram is:

$$[\![\langle p_3, p_{3.1} \rangle]\!] : E_1 \rightarrow \square(E_{62} \times E_{55})$$

The qualifiers of DC.Description are:

- Abstract
- Notes
- Contents

"Type" applies to all these qualifiers.

5.2.3.8 Mapping of DC.Publisher

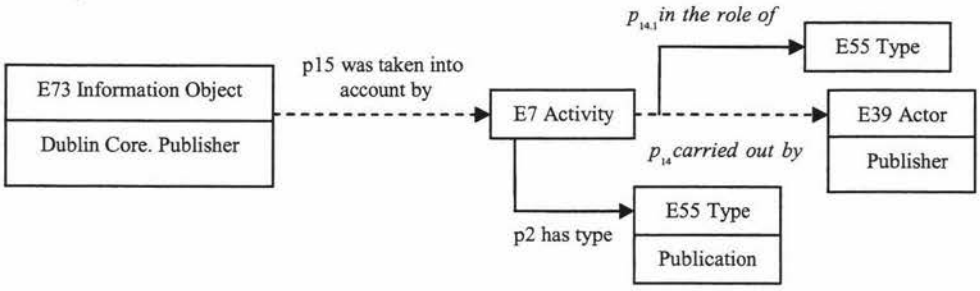
DC.Publisher refers to the entity responsible for making the resource available. There are three interpretations for publishing, and the mapping of Publisher is based on these interpretations (Doerr, 2000):

- The resource itself is the publication

  In this case, publishing is regarded as the publication with an ISBN number or the copy is accessible from the Internet.

  In this case, DC.Publisher is represented by E_{73}*Information Object* :

  Mapping diagram:



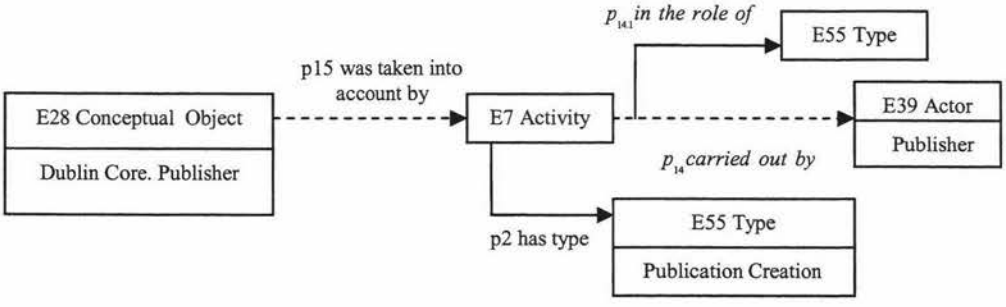
  Notation:

$$[\![p_{15} \langle p_2, [\![\langle p_{14}, p_{14.1} \rangle]\!] \rangle]\!] : E_{73} \rightarrow \square \left(E_7 \times \left(E_{55} \times \square \left(E_{55} \times E_{39} \right) \right) \right)$$

- Publishing is regarded as serious action of creating publishing

  In this case, DC.Publisher is represented b E_{28}*Conceptual Object* . There is a possibility of multiple "publication creation". In such a case, the property "in the role of" on top of the property "carried out by" could be used to distinguish the different roles.

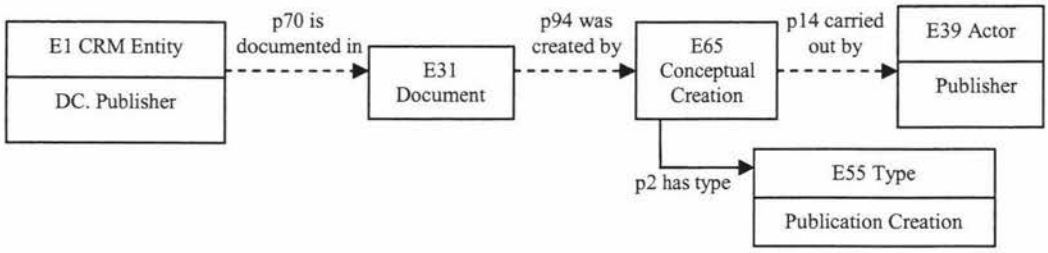
  Mapping diagram:



  Notation:

$$[\![p_{15} \langle p_2, [\![\langle p_{14}, p_{14.1} \rangle]\!] \rangle]\!] : E_{28} \rightarrow \square \left(E_7 \times \left(E_{55} \times \square \left(E_{55} \times E_{39} \right) \right) \right)$$

- Other forms of publication. In this case, DC.Publisher is restricted by related DCT1 element, such as physical object, party, event, so there is DC [DCT1 physical object, party, event].



Notation:

$$[\![p_{70}[\![p_{94}\langle p_2,[\![p_{14}]\!]\rangle]\!]]\!]: E_1 \rightarrow \square\left(E_{31}\times\square\left(E_{65}\times\left(E_{55}\times\square E_{39}\right)\right)\right).$$

5.2.3.9 *Mapping of DC.Date*

DC.Date represents a date of an event in the lifecycle of the resource. Mapping of DC.Date is associated with the DC.Date qualifier, the DC.Date qualifiers are:

- Created
- Modifier
- Issue
- Available
- Valid

Before starting the mapping of these qualifiers, the mapping below applies to all the cases of DC.Date:

- $DC.Date = E_{52}Time-Span$

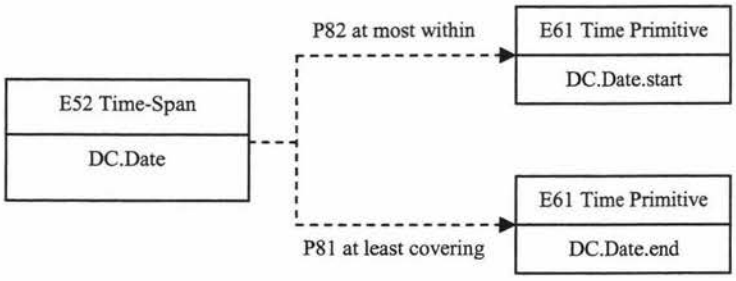
  "start" and "end" of date are regarded as an interval within which something has happened or a duration of a certain process:

  $DC.start = E_{61}Time\ Primitive$

  $DC.end = E_{61}Time\ Primitive$

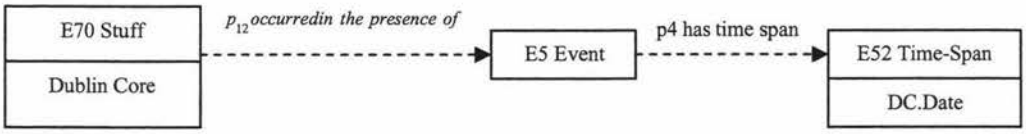
  Following are the mappings of Date.start and Date.end, we can choose either one of them or both of them as the property to Time-Span:

Mapping diagram:



Notation: $[\![p_{82}, p_{81}]\!] : E_{52} \rightarrow \square(E_{61}, E_{61})$

- When the DC.Date is used without the restriction of the qualifier, Doerr (2000) introduces the mapping DC.Date as $E_{70}Stuff$ when there's no specified event:

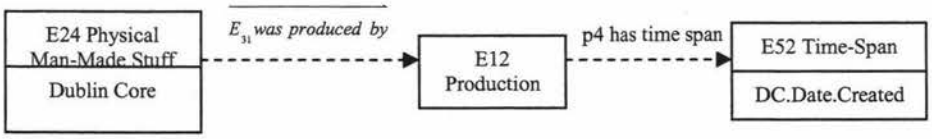


Notation: $[\![p_{12} [\![p_4]\!]]\!] : E_{70} \rightarrow \square(E_5 \times \square E_{52})$

- When the qualifier of DC.Date is "Created":

  According to Doerr (2000), the mapping of qualifier "Created" for DC.Date splits into two parts:

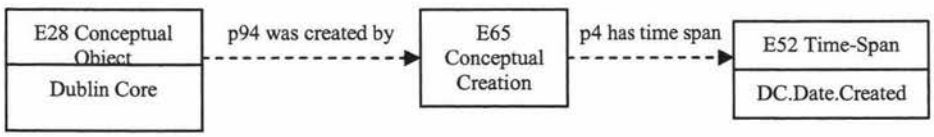
  Mapping diagram: When DC.Date is restricted by CRM entity $E_{24}Physical\ Man-Made\ Stuff$:



Notation:

$[\![\overline{p_{31}} [\![p_{41}]\!]]\!] : E_{52} \rightarrow \square(E_{12} \times \square E_{52})$

Mapping diagram: When DC.Date is associated with CRM entity $E_{28}Conceptual\ Object$:

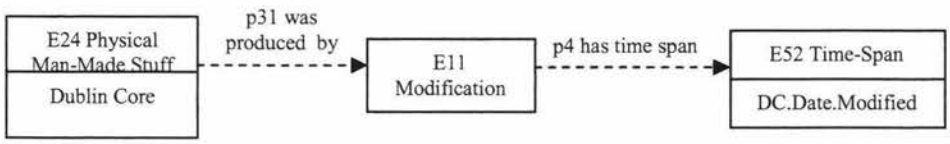


Notation:

$[\![p_{94}[\![p_4]\!]]\!]: E_{28} \rightarrow \square(E_{65} \times \square E_{52})$

- When the qualifier of DC.Date is "Modified".

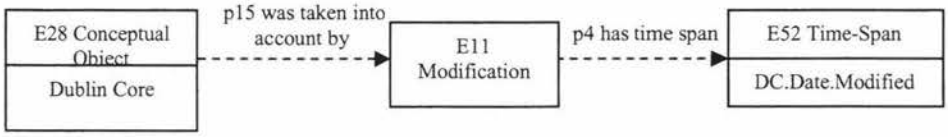
  Mapping diagram: When DC.Date is associated with CRM entity $E_{24} Physical\ Man-Made\ Stuff$:



Notation:

$[\![p_{31}[\![p_4]\!]]\!]: E_{24} \rightarrow \square(E_{11} \times \square E_{52})$

Mapping diagram: When DC.Date is associated with CRM entity $E_{28} Conceptual\ Object$:



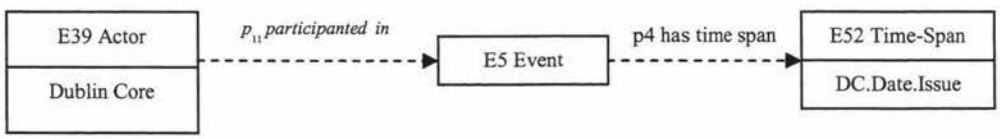
Notation:

$[\![p_{15}[\![p_4]\!]]\!]: E_{28} \rightarrow \square(E_{11} \times \square E_{52})$

- When the qualifier of DC.Date is "Issued"

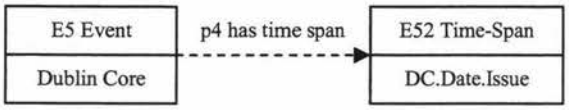
  "Issued" is related to the "Publication" event, in this case, DC.Date.Issued is associated with either $E_{39} Actor$ or $E_5 Event$.

  Mapping diagram: When the DC.Date is restricted by $E_{39} Actor$, the mapping diagram is:



Notation: $[\![p_{11}[\![p_4]\!]]\!]: E_{39} \rightarrow \square(E_{65} \times \square E_{52})$

Mapping diagram: When the DC.Date is restricted by $E_5 Event$, the mapping diagram is:



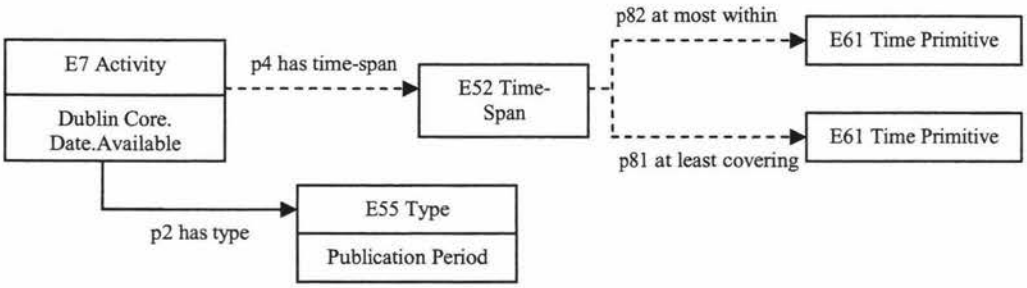
Notation:

$[\![p_4]\!]: E_5 \rightarrow \square E_{52}$

- Qualifier "Available"

According to Doerr (2000), DC.Date.Available can be related to $E_7 Activity$, which has the type of 'Publication period'; and 'Time-span' related to begin and end.

Mapping diagram:



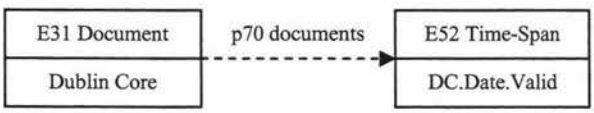
Notation:

$\langle p_2, [\![p_4 [\![p_{82}, p_{81}]\!]]\!] \rangle : E_7 \rightarrow E_{55} \times \square (E_{52} \times \square (E_{61}, E_{61}))$

- Qualifier "Valid"

Suggested by Doerr (2000), DC.Date.Valid is associated with $E_{31} Document$.

Mapping diagram:

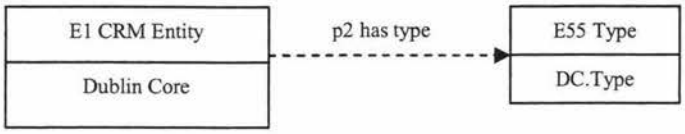


The notation of it is: $[\![p_{70}]\!]: E_{31} \rightarrow \square E_{52}$

5.2.3.10 Mapping of DC.Type

DC.Type represents the nature or genre of the content of the resource. DC.Type maps to $E_{55} Type$, this leads to the mapping of DC.Type.

Mapping diagram:

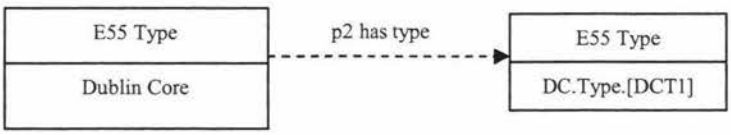


Notation:

$[\![p_2]\!]: E_1 \rightarrow \square E_{55}$.

The qualifier for DC.Type is [DCT1], DC.Type.[DCT1] maps to $E_{55}Type$.

Mapping diagram:



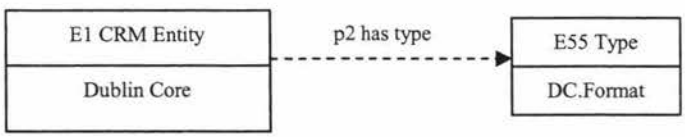
Notation:

$[\![p_2]\!]: E_{55} \rightarrow \square E_{55}$

5.2.3.11 Mapping of DC.Format

DC.Format refers to the physical or digital manifestation of the resource. Based on Doerr's (2000) interpretation of DC.Format is a type. Similar with the mapping of DC.Type: $DC.Format = E_{55}Type$:

Mapping diagram:



Notation:

$[\![p_2]\!]: E_1 \rightarrow \square E_{55}$

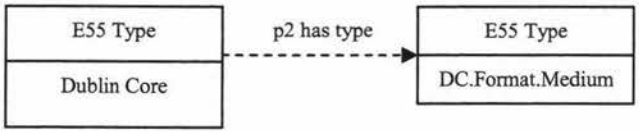
There are three qualifiers for DC.Format, they are:

- Medium: the material or physical carrier of the resource),
- IMT: the Internet media type of the resource)
- Extent: the size or duration of the resource).

Doerr (2000) maps both Medium and IMT to $E_{55}Type$:

- $DC.Format.Medium = E_{55}Type$

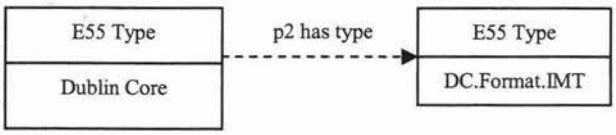
  Mapping diagram:



  Notation:

  $[\![p_2]\!]: E_{55} \rightarrow \square E_{55}$

- $DC.Format.IMT = E_{55}Type$

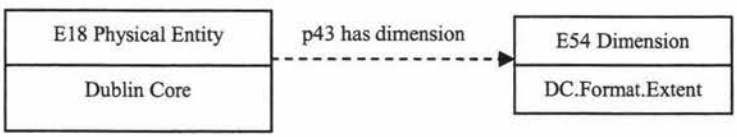


  Notation:

  $[\![p_2]\!]: E_{55} \rightarrow \square E_{55}$

- The mapping of qualifier Extent depends on the resource type. As the CRM covers dimensions for physical entities, therefore, Dublin Core maps to $E_{18}Physical\ Entity$ and DC.Format.Extent maps to $E_{54}Dimension$:

  Mapping diagram:

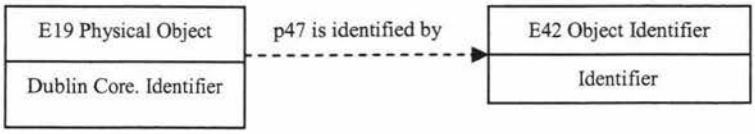


  Notation:

  $[\![p_{43}]\!]: E_{18} \rightarrow \square E_{54}$

5.2.3.12 Mapping of DC.Identifier

DC.Identifier refers to an unambiguous reference to the resource within a given context. According to Doerr (2000), an identifier refers to the unique number represent the Dublin Core resource. In CRM, the identifiers would be treated as Appellations. In this case, the DC.Identifier is associated with $E_{19}Physical\ Object$, and DC.Identifier maps to $E_{42}Object\ Identifier$, see the mapping diagram below:



Notation for the above diagram:

$[\![p_{47}]\!]: E_{19} \rightarrow \square E_{42}$

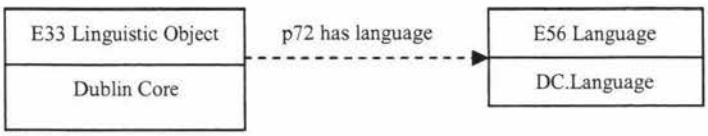
5.2.3.13 Mapping of DC.Source

DC.Source is a reference to a resource from which the present resource is derived. Mapping of DC.Source is the same with the mapping of DC.Relation, refer to 5.2.3.14 for mapping details.

5.2.3.14 Mapping of DC.Language

DC.Language is a language of the intellectual content of the resource. Language can be treated as $E_{56}Language$ in CRM, and in this case, Dublin Core is associated with $E_{33}Linguistic\ Object$.

Mapping diagram:



Notation:

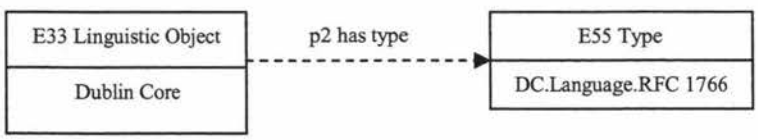
$[\![p_{72}]\!]: E_{33} \rightarrow \square E_{56}$.

There are two qualifiers for Language:

- ISO 639-2: codes for the representation of names of languages
- RFC 1766: specifies a two letter code taken from ISO 639, followed optionally by a two letter country code taken from ISO 3166.

When DC.Language related to these two qualifiers, DC.Langage can be treated as the Type linked by the property "has type" to the entity $E_{33}Linguistic\ Object$.

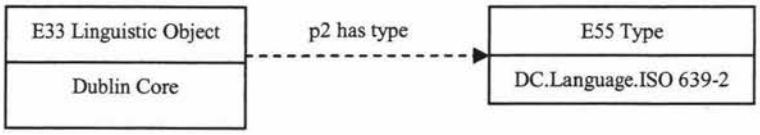
- Mapping diagram: When the qualifier of DC.Language is RFC1766



Notation:

$[\![p_2]\!]: E_{33} \rightarrow \square E_{55}$

- Mapping diagram:When the qualifier of DC.Language is ISO 639-2



Notation:

$[\![p_2]\!]: E_{33} \rightarrow \square E_{55}$.

5.2.3.15 Mapping of DC.Relation

DC.Relation refers to a reference to a related resource. Doerr (2000) points out that in CRM, series relations (such as properties: "is composed of", "is referred to") has to be expressed in a clear way through the use of events, which is similar to DC.Relation. During the mapping of DC.Relation, Doerr (2000) suggests that mapping of DC.Relation should be associated with its qualifiers. The Dublin Core qualifiers for DC.Realtion are:

- HasPart: The described resource includes the referenced resource either physically or logically.
- IsPartOf: The described resource is a physical or logical part of the referenced resource.
- References: The described resource references, cites, or otherwise points to the referenced resource.
- HasReferencedBy: The described resource is referenced, cited, or otherwise pointed to by the referenced resource.
- IsVersionOf: The described resource is a version, edition, or adaptation of the referenced resource. Changes in version imply substantive changes in content rather than differences in format.
- HasVersion: The described resource has a version, edition, or adaptation, namely, the referenced resource.
- IsFormatOf: The described resource is the same intellectual content of the referenced resource, but presented in another format.
- HasFormat: The described resource pre-existed the referenced resource, which is essentially the same intellectual content presented in another format.

- IsRequiredBy: The described resource is required by the referenced resource, either physically or logically.
- Requires: The described resource supplants, displaces, or supersedes the referenced resource.

▪ Mapping of DC.Relation.HasPart and DC.Relation.IsPartOf.

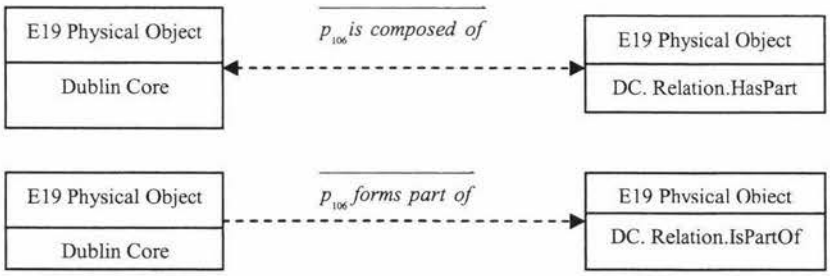
In this mapping, DC.Relation.HasPart and DC.Relation.IsPartOf can be referred to theses properties:

" p_{106}*is compsed of* (*forms part of*) "," p_{46}*is compsed of* (*forms part of*) "," p_{9}*consists of* (*forms part of*) " and" p_{88}*consists of* (*forms part of*) " that associated with CRM entities:

E_{19}*Physical Object* , E_{28}*Conceptual Object* , E_{53}*Place* , E_{4}*Period* and E_{74}*Group* .

- When DC.Relation.HasPart and DC.Relation.IsPartOf are associated with E_{19}*Physical Object* :

  Mapping diagram:

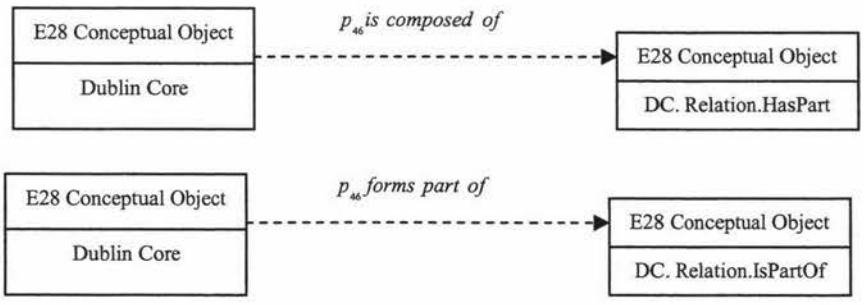


  Notation for both diagrams:

  $[\![\overline{p_{106}}]\!] : E_{19} \rightarrow \Box E_{19}$

- When DC.Relation.HasPart and DC.Relation.IsPartOf are associated with E_{28}*Conceptual Object* :

  Mapping diagram:

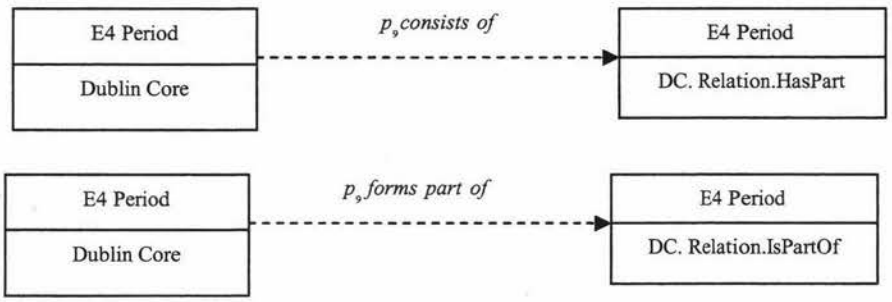


Notation for both diagrams:

$[\![p_{46}]\!]: E_{28} \rightarrow \square E_{28}$

- When DC.Relation.HasPart and DC.Relation.IsPartOf are associated with $E_4 Period$:

Mapping diagram:

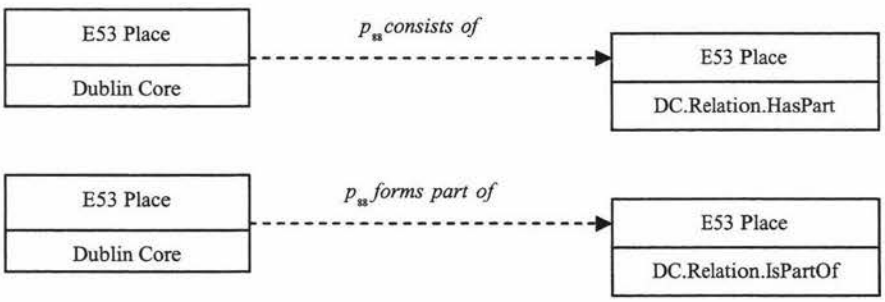


Notation for both diagrams:

$[\![p_9]\!]: E_4 \rightarrow \square E_4$

- When DC.Relation.HasPart and DC.Relation.IsPartOf are associated with $E_{53} Place$:

Mapping diagram:

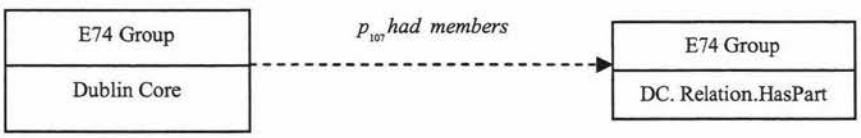


Notation for both diagrams:

$[\![p_{88}]\!]: E_{53} \rightarrow \square E_{53}$

- When DC.Relation.HasPart is associated with $E_{74} Group$:

Mapping diagram:

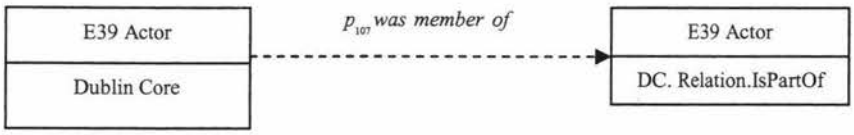


Notation:

$[\![p_{107}]\!]: E_{74} \rightarrow \square E_{74}$

- When DC.Relation.HasPart is associated with $E_{39} Actor$:

Mapping diagram:



Notation:

$[\![p_{107}]\!]: E_{39} \rightarrow \square E_{39}$

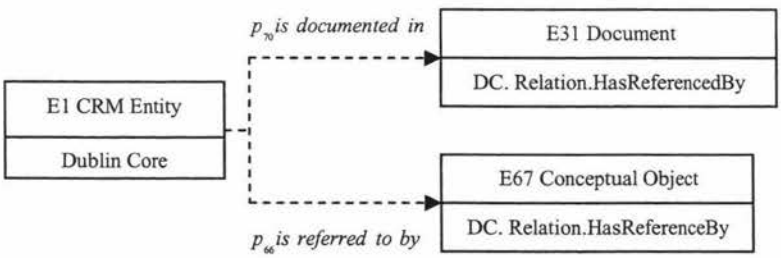
- Mapping of DC.Relation.References and DC.Relation.HasReferencedBy

  Doerr (2000) suggests that three cases can be applied for the mapping of "DC.Relation.References" and "DC.Relation.HasReferencedBy". These three cases are: the reference of valid information about documentation; the related document made in conceptual object; and the description on the physical objects.

  - In the first case, when DC.Relation.HasReferencedBy and DC.Relation.References are refer to the reference of valid information about documentation, see the following mapping for details.

    DC.Relation.HasReferencedBy refers to the document is documented in or is referred to by another document, it can be regarded as the properties " p_{70}*is documented in* " and " p_{66}*is referred to* " that DC.Relation.HasReferencedBy is associated with E_1*CRM Entity* :

    Mapping diagram:

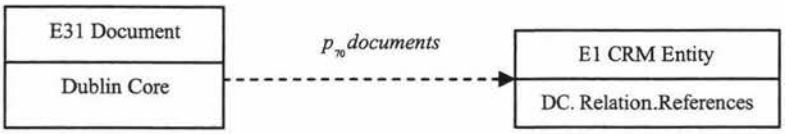


    Notation:

    $[\![p_{70}, p_{70}]\!]: E_1 \rightarrow \square\left(E_{31}, E_{67}\right).$

    When DC.Relation.References refers to document the references, then DC.Relation.References can be regarded as the properties " p_{70}*documents* " that DC.Relation.References is associated with E_{31}*Document* :

Mapping diagram:

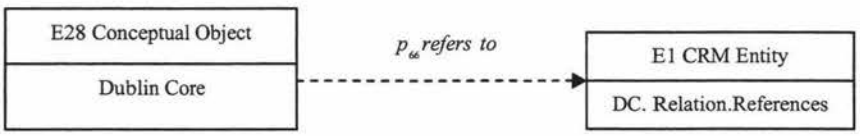


Notation:

$[\![p_{70}]\!]: E_{31} \rightarrow \square E_1$

- In the second case, when the DC.Relation.References is related to document made in conceptual object, DC.Relation.References can be regarded as the property " p_{66}*refers to* " that associated with E_{28}*Conceptual Object* :

  Mapping diagram:



  Notation:

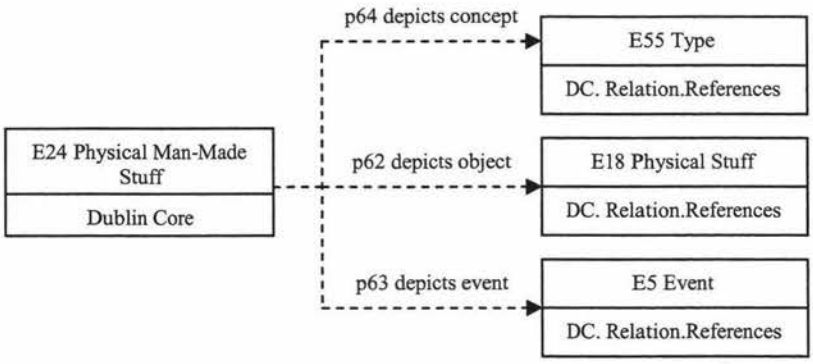
  $[\![p_{66}]\!]: E_{28} \rightarrow \square E_1$

- In the third case, when the DC.Relation.References and DC.Relation.HasReferencedBy are related to the depictions on/by the shape of physical object, see the mapping details below:

  When DC.Relation.References refers to the 'Physical Man-Made Stuff' depicts a concept; or to depict an object; or to depict an event, DC.Relation.References can be regarded as the properties:

  " p_{64}*depicts concept* ", " p_{62}*depicts object* " and " p_{63}*depicts event* ":

  Mapping diagram:

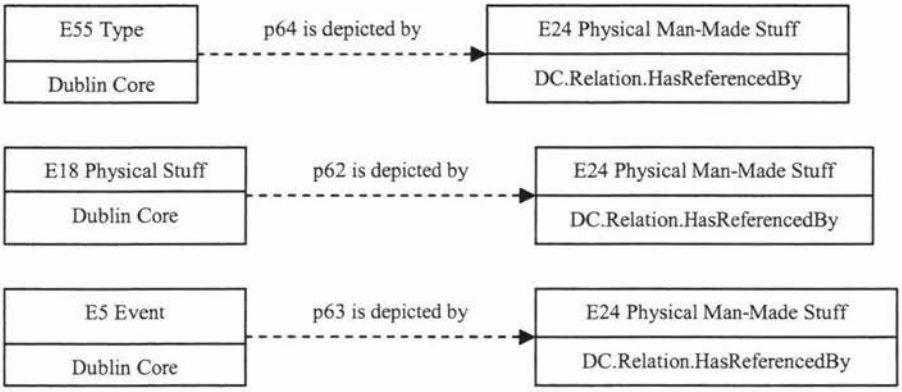


  Notation:

  $[\![p_{64}, p_{62}, p_{63}]\!]: E_{24} \rightarrow \square\left(E_{55}, E_{18}, E_{63}\right).$

- When DC.Relation.HasReferencedBy refers to the type, the physical stuff and the event is depicted by the physical man-made stuff, DC.Relation.HasReferencedBy can be regarded as properties: "p_{64}*is depicted by*", "p_{62}*is depicted by*" and "p_{63}*is depicted by*":

  Mapping diagram:



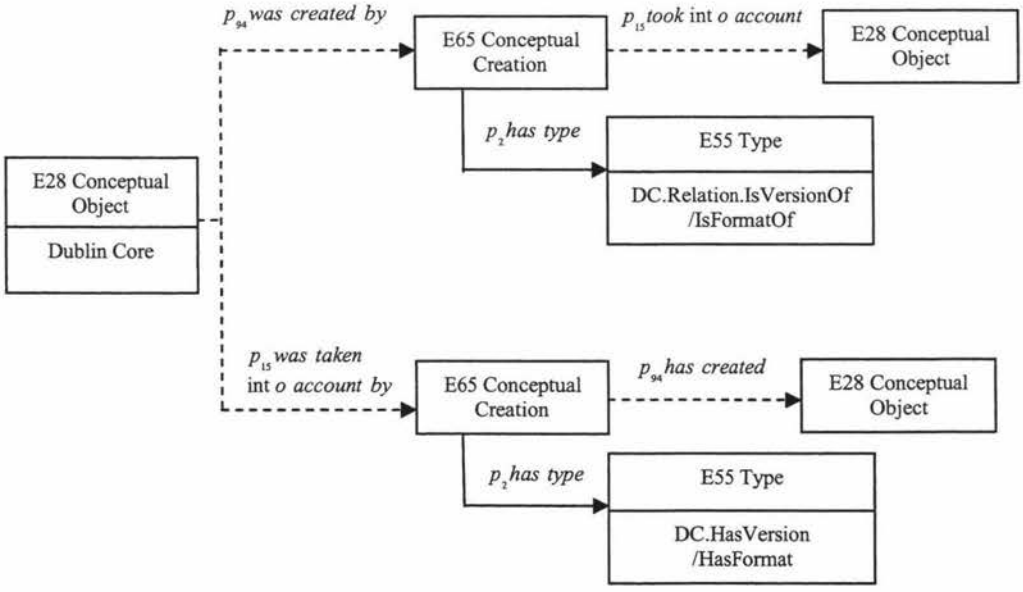
  These diagrams lead to the notation:

  $[\![p_{64}]\!]: E_{55} \rightarrow \square E_{24}$, $[\![p_{62}]\!]: E_{18} \rightarrow \square E_{24}$ and $[\![p_{63}]\!]: E_5 \rightarrow \square E_{24}$.

- Mapping of "IsVersionOf/HasVersion" and "IsFormatOf/HasFormat":

  Doerr (2000) suggests that the mapping of "IsVersionOf/HasVersion" and "IsFormatOf/HasFormat": are related to the properties "p_{94}*was created by*" and "p_{15}*was taken into account by*" that are associated with the creation of the respective resource:

  Mapping diagram:



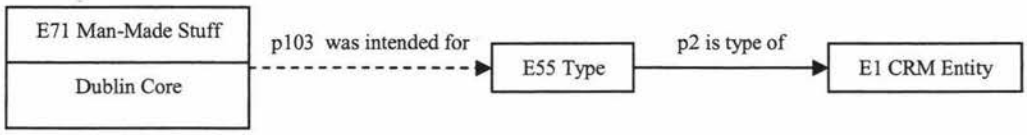
Notation:

$$[\![p_{94}\langle[\![p_{15}]\!], p_2\rangle, p_{15}\langle[\![p_{94}]\!], p_2\rangle]\!]: E_{28} \rightarrow \square\left(E_{65} \times \left(\square E_{28} \times E_{55}\right), E_{65} \times \left(\square E_{55} \times E_{28}\right)\right).$$

- Mapping of "IsRequiredBy"

  "IsRequiredBy" is regarded as the CRM property "was intended for" which associated with 'Man-Made Stuff' and its related Type. Doerr (2000) points out that type of the requiring resource should be match with the type of things the resource came from:

  Mapping diagram:



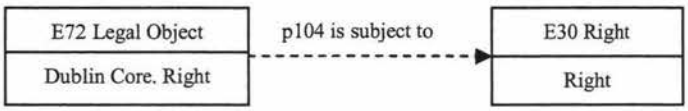
  Notation:

$$[\![p_{103}\langle p_2\rangle]\!]: E_{71} \rightarrow \square\left(E_{55} \times E_1\right).$$

5.2.3.16 *Mapping of DC.Right*

DC.Right represents information about rights held in and over the resource. When the resource is related to copy rights, DC.Right can be mapped to $E_{30}Right$ that associate with the entity $E_{72}Legal\ Object$:

Mapping diagram:



Notation:

$$[\![p_{104}]\!]: E_{72} \rightarrow \square E_{30}.$$

5.3 Mapping of the AMICO data model to the CIDOC CRM version 2.3

5.3.1 Introduction to AMICO

The Art Museum Image Consortium (AMICO, http://www.amico.org/) was formed in 1997. It is a non-profit consortium. The consortium has large numbers of art collections; they are estimated to have over 10,000 works of arts. The AMICO library is a licensed educational resource available under subscription to universities and

colleges, public libraries, elementary and secondary schools and museums. It has more than 30 members, and the AMICO library is the gathering for digital multimedia art works.

Based on AMICO Data Specification (2002), each art work in AMICO data is documented by: a category record, multimedia files associated with the art works and a metadata record.

5.3.1.1 Data dictionary format

To understand the structure of AMICO data structure, refer to Appendix 3 for the copy of the AMICO Data Specification: Data Dictionary Version 1.3 (2002). Following is a description of the Data Dictionary Format (AMICO – Data Dictionary Version 1.3, 2002):

- ***TAG:*** three letter prefix that identifies an AMICO data field, such as "AID" which represents the field name: "AMICO identifier".
- ***Field Name***: the full name of the field.
- ***Core***: fields that are required to be present in an AMICO Library Record.
- ***Repeat***: whether a field can occur more than once in a record, such as AID is the unique identifier in AMICO record, so in this case, AID cannot be repeated.
- ***Group***: if a field belongs to a group (or is group tag), such as OTG is a group tag for the group of fields contain title/name of the art work, and field like "Title-Type" belongs to OTG.
- ***Definition/Guidelines***: what information is recorded in a field and how it is structured.
- ***Examples***: samples of the kinds of data that will be found in the field.
- ***Version:*** the version of the Data Dictionary where the field first appeared.

In the following sections, descriptions of AMICO category record fields are quoted from AMICO Data Specification: Data Dictionary Version 1.3 (2002).

5.3.1.2 Category record

The category record is based on the Categories for the Description of Works of Art; it is a product of the Art Information Task Force (AITF). Baca and Harpring, (2000)

describe the category record as a template for describing works of art, architecture, groups of objects, and visual and textual information. Baca and Harpring describe the content of an art work by establishing the conceptual framework that enables it to access information about the object and its images. The fact that the category record describes and identifies the object, allows the information to fit into diverse systems and makes the information more accessible. They consider the use of the framework will help retain the integrity of museum data and contribute to its longevity. It will also assist the inevitable migration of the data to new systems as informational technology continues to evolve.

In the AMICO category record, record fields are categorised as:

- Unique identification
- What is it?
- Who made it?
- When was it made?
- Where was it made?
- What is it about?
- What does it mean?
- Who showed it?
- Who owns it?
- What is it related to?
- Who documented it?

5.3.1.3 *Associated multimedia files*

Each art work in the collection system must contribute at least one image of the whole art work. All associated media files, text, image, multimedia, follow the same naming and linking conventions. (AMICO Related Image and Multimedia Files Specification Version 1.2, 2002) In addition, each image or other media file will be accompanied by a separate structured text-based metadata record, and is also referenced by an entry in the Related-Multimedia group field of the AMICO Catalogue.

5.3.1.4 *Metadata record*

The AMICO metadata record is used to record the associated multimedia files as mentioned above. The Media Metadata record is based on the Dublin Core record.

5.3.2 Mapping formalism

The mapping in this thesis is based on Doerr.'s first mapping in 2000 (Doerr, 2000), and it is based on CRM version 3.0 and the AMICO data dictionary version 1.2.

5.3.2.1 Mapping scheme

The mapping formation is same with the validation applied in the previous sections. Refer to section 5.1.2.2 for details.

5.3.3 Mapping AMICO data to CRM

Doerr (2001a) states that mapping AMICO to the CRM takes place at two semantic levels. All the AMICO dataset should be treated as an instance of $E_{31}Document$. In addition, the contents of the AMICO dataset about an object maps to its corresponding real-world entities in the CRM to which it refers and relates.

5.3.3.1 Mapping for "Unique Identification"

As mentioned above, AID (AMICO Identifier), the complete dataset of AMICO object can be regarded as $E_{31}Document$. Hence,

$AID = E_{31}Document$

The AMICO record itself can be treated as an object in CRM. The CR (Category Record), itself maps to E_1 CRM Entity. This is expressed as:

$CR = E_1$ CRM Entity. The correspondence relationship between AID and CR can be expressed as:

$AID.CR = E_{31}Document : (p_{70}Document) : E_1\ CRM\ Entity$

Transferring to category theory notation as:

$\langle p_{70} \rangle : E_{31} \rightarrow E_1$.

5.3.3.2 Mapping for "What is it?"

- OTY (Object-Type) maps to $E_{55}Type$, the subclass of $E_{22}Man-Made\ Object$, so there is: $OTY = E_{55}Type$, the corresponding relationship between OTY and CR becomes:

  $CR.OTY = E_{22}Man-Made\ Object : p_2 has\ type : E_{55}Type$

  Notation:

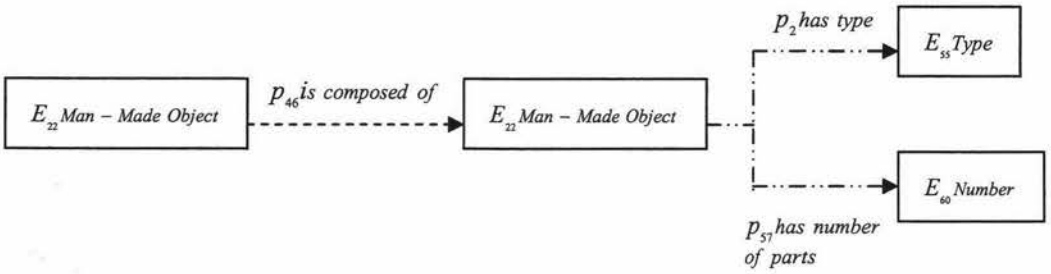
  $[\![p_2]\!] : E_{22} \rightarrow E_{55}$

- OPP (Object-Parts/Pieces) can be treated as object that decomposed into certain types and the object has certain numbers of parts; this can be represented as:

$OPP = E_{55}Type$ and (or) $OPP = E_{60}Number$

The corresponding relationship between OPP and CR is expressed in the following diagram:

Mapping diagram:



Notation:

$[\![p_{46}[p_2, p_{57}]]\!] : E_{22} \rightarrow \square(E_{22} \times (E_{55} \oplus E_{60}))$

- CLG (Classification Group) can be treated as $E_{55}Type$ of the object in CRM: $CLG = E_{55}Type$, its correspondence relationship with CR becomes: $CR.CLG = E_{22}Man - Made\ Object : p_2 has\ type : E_{55}Type$.

  CLT, CLS, are the fields under the CLG group:

  - CLT (Classification-Term) maps to the type of the object: $CLT = E_{55}Type$, with the link between CLT and CR, there is: $CR.CLT = E_{22}Man - Made\ Object : p_2 has\ type : E_{55}Type$.

    Notation:

    $[\![p_2]\!] : E_{22} \rightarrow E_{55}$.

  - CLS (Classification-Scheme) is a classification scheme from which a term was chosen, it can be mapped as $E_{32}Authority\ Document$ in CRM, and this can be expressed as: $CLS = E_{32}Authority\ Document$, when related with the link of CR, there is:

    $CR.CLS = E_{22}Man - Made\ Object : p_{71} is\ part\ of : E_{32}Authority\ Document$.

    Notation:

    $[\![p_{71}]\!] : E_{22} \rightarrow E_{32}$.

5.3.3.3 What is it called?

- OTG (Object-Title/Name Group) corresponds to the title/name of the object in CRM. It maps to $E_{35}Title$ in CRM, its correspondence relationship with CR becomes: $CR.OTG = E_{22}Man-Made\ Object : p_{102}has\ title : E_{35}Title$.

  Notation:

  $[\![p_{102}]\!] : E_{22} \rightarrow E_{35}$.

  OTN and OTT are two fields belong to the OTG group:

  - OTN (Object-Title-Name) maps as the title of the object, this can be expressed as: $OTN = E_{35}Title$, the correspondence relationship with CR is: $CR.OTN = E_{22}Man-Made\ Object : p_{102}has\ title : E_{35}Title$.

    Notation:

    $[\![p_{102}]\!] : E_{22} \rightarrow E_{35}$

  - OTT (Object-Title-Type) can be mapped to the type of the object: $OTT = E_{55}Type$, the correspondence relationship with CR can be expressed as: $CR.OTT = E_{22}Man-Made\ Object : p_{2}has\ type : E_{55}Type$.

    Notation:

    $[\![p_{2}]\!] : E_{22} \rightarrow E_{55}$.

- OST (State) is the data of the unique process that created the multiple, so OST can be treated as $E_{11}Modification$, hence, $OST = E_{11}Modification$. When OST relates with CR, there is the relationship:

  $CR.OST = E_{22}Man-Made\ Object : p_{31}was\ produced\ by : E_{11}Modification$.

  Notation:

  $[\![p_{31}]\!] : E_{22} \rightarrow E_{11}$.

  Or OST can be treated as the event note (or extension) comes along with the object, this leads to: $OST = E_{62}String$, its relation with CR can be written as:

  $CR.OST = E_{22}Man-Made\ Object : p_{3}has\ note : E_{62}String$.

  Notation:

  $[\![p_{3}]\!] : E_{22} \rightarrow E_{62}$.

- OEN (Edition) is for works produced in multiples, the edition of this particular example. It can be mapped as $OEN = E_{29} Design\ or\ Procedure$, its correspondence relationship with CR is:

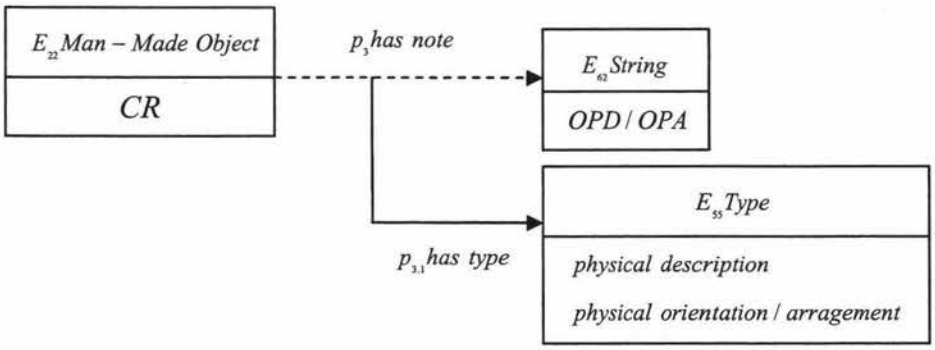
  $CR.OEN = E_{22} Man-Made\ Object : p_{11} was\ produced\ by :$
  $E_{12} Production : p_{33} used\ specific\ technique : E_{29} Design\ or\ Procedure$

  Notation: $[\![\langle p_{33}, p_{11} \rangle]\!] : E_{22} \rightarrow \square(E_{12} \times E_{29})$.

5.3.3.4 What does it look like?

- OPD (Physical Description) and OPA (Physical Orientation/Arrangement) are all associated with the description of the artwork. OPD is the description for the physical appearance of the component of the object, it can be mapped as $E_{62} String$: $OPD = E_{62} String$. OPA is used to describe the orientation of the artwork, such as the description of how to assemble the artwork. It can be treated as $E_{62} String$ in CRM. Mapping of OPD and OPA are expressed in the following diagram:

  Mapping diagram:



  Notation:

  $[\![p_3, p_{3.1}]\!] : E_{22} \rightarrow \square(E_{62} \times E_{55})$

- MET (Measurements-Text) can be interpreted as the text associated with the work's measurement, same with OPD and OPA, It can be treated as $E_{62} String$. Its relationship with CR can be expressed as:

  $CR.MET = E_{22} Man-Made\ Object : p_3 has\ note : E_{62} String$.

  Notation:

  $[\![p_3]\!] : E_{22} \rightarrow E_{62}$.

- MEG (Measurements Group) is used to group the fields recording the measurement of the object. These fields include MCM, MED, MDV and MDU.

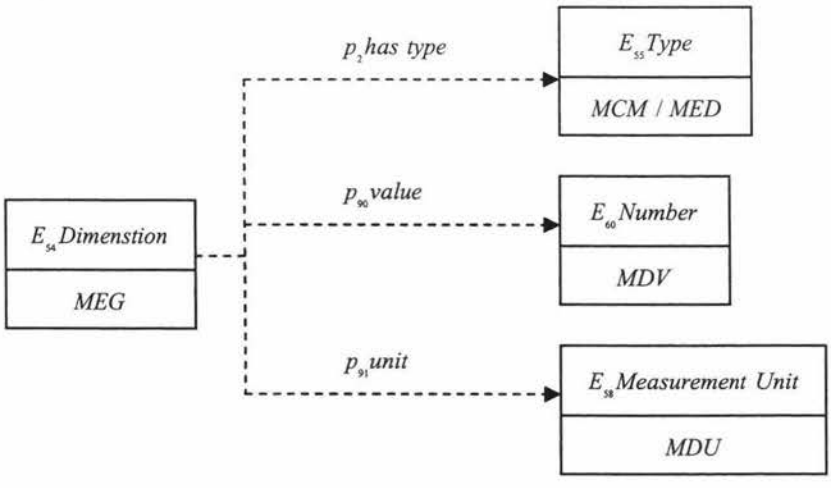
  MEG corresponds to $E_{54}Dimension$ in CRM: $MEG = E_{54}Dimension$. Its correspondent relationship with CR can be expressed as: $CR.MEG = E_{22}Man-Made\ Object : p_{43}has\ dimension : E_{54}Dimension$.

  Notation:

  $[\![p_{43}]\!] : E_{22} \rightarrow E_{54}$

  - MCM (Measurement-Component-Measured) is the term used to indicate what was measured, such as the frame, the lid...etc. It can be mapped to $E_{55}Type$ in CRM. MED (Measurement-Dimension) is the term used to indicate the measurement taken, such as: height, width…etc. It maps to $E_{55}Type$ in CRM.
  - MDV (Measurement-Dimension-Value) represents the value of the measurement. It can be mapped as $E_{60}Number$.
  - MDU (Meaurement-Dimension-Unit) represent the unit in which the measurement is taken. It maps to $E_{58}Measurement\ Unit$ in CRM. The following diagram describes mapping of MEG:

  Mapping diagram:



  Notation:

  $[\![p_2, p_{90}, p_{91}]\!] : E_{54} \rightarrow \square(E_{55}, E_{60}, E_{58})$.

- OMG (Materials and Techniques Group) is used to present the group associated with the production process. The subfields of OMG include: OMD, OMT, OMM and OMS.

OMG can be mapped as $E_{12}Production : OMG = E_{12}Production$, when it link with CR, its relationship can be described as: $CR.OMG = E_{22}Man-Made\ Object : p_{31}was\ produced\ by : E_{12}$ Pr$oduction$.

Notation: $[\![p_{31}]\!] : E_{22} \rightarrow E_{12}$.

Also suggested by Theodoridou and Doerr (2001), if a specific production process is involved, and there is the aspect of "material", OMG corresponds to $E_{29}Design\ or$ Pr$ocedure$, its relationship with CR can be expressed as:

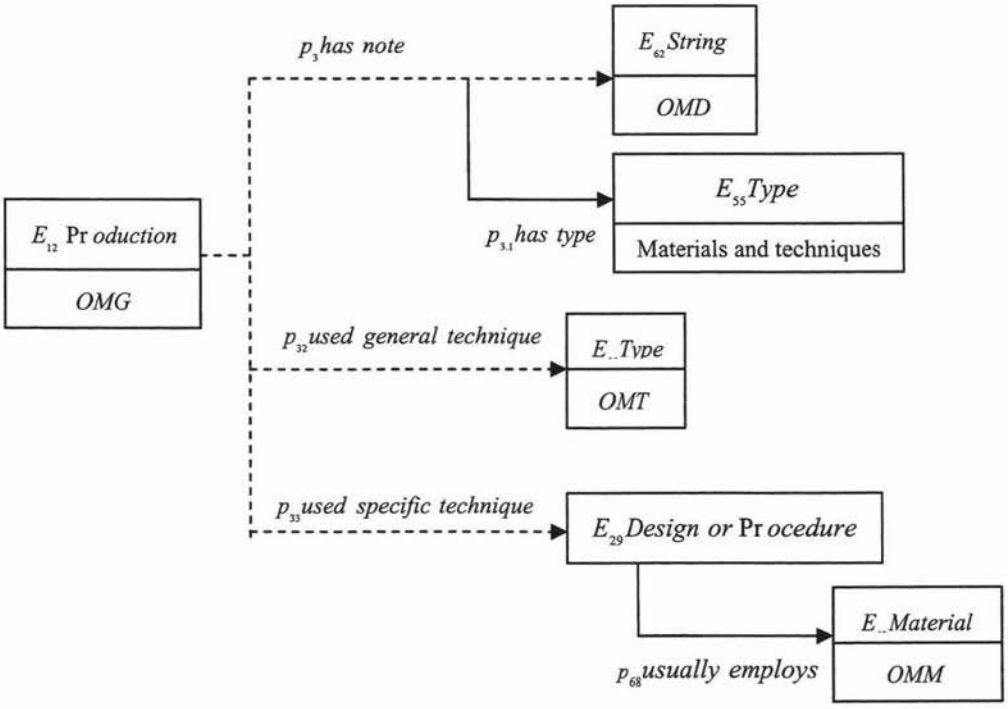
$CR.OMG = E_{22}Man-Made\ Object : p_{31}was\ produced\ by :$
$E_{11}Modification : p_{33}used\ specific\ technique : E_{29}Design\ or$ Pr$ocedure$

Notation:

$[\![p_{31}, p_{33}]\!] : E_{22} \rightarrow \square(E_{11} \times E_{29})$.

- OMD (Materials and Techniques-Description) is the text description of the techniques and material used to create the work. It can be mapped as $E_{62}String$.
- OMT (Materials and Techniques-Process/Technique-Term) is the term used to describe the processes and techniques used to create the art work. It maps to $E_{55}Type$.
- OMM (Materials and Techniques-Materials-Term) is the term used to describe the materials to create the art work. It maps to $E_{57}Material$.

The following diagram represents the mapping of MEG:



Notation:

$[\![\langle p_3, p_{3.1}\rangle, p_{32}, p_{33}]\!] : E_{12} \rightarrow \square(E_{62} \times E_{55}, E_{55}, E_{29} \times E_{57})$.

Besides, OMM is able to link with CR:

$CR.OMM = E_{22}Man-Made\ Object : p_{45}consist\ of : E_{57}Material$.

Notation: $[\![p_{45}]\!] : E_{22} \rightarrow E_{57}$

- OMS (Materials and Techniques-Support) refers to single terms that index the support on which the work was created. Doerr (2001a) suggests to map OMS to both $E_{55}Type$ and $E_{57}Material$, and the relationship between OMS and CR can be described as:

  $CR.OMS = E_{22}Man-Made\ Object : p_{46}is\ composed\ of : E_{18}Physical\ Stuff :$
  $p_{31}was\ produced\ by : E_{11}Modification : p_{32}used\ general\ technique : E_{55}Type$

  and $CR.OMS = E_{22}Man-Made\ Object : p_{46}is\ composed\ of :$
  $E_{18}Physical\ Stuff : p_{45}consist\ of : E_{57}Material$.

  Notation:

  $[\![\langle p_{46}, p_{31}, p_{32}\rangle]\!] : E_{22} \rightarrow \square(E_{18} \times \square(E_{11} \times \square E_{55}))$

  and $[\![\langle p_{46}, p_{45}\rangle]\!] : E_{22} \rightarrow \square(E_{57} \times \square E_{18})$.

- OIN (Inscriptions and/or Marks) is a text description for any inscriptions or marks on the artwork. It corresponds to $E_{37}Mark$, and it can be mapped as $E_{62}String$ in CRM. Its relationship with CR can be expressed as:

  $CR.OIN = E_{22}Man-Made\ Object : \overline{p_{65}shows\ visual\ item} :$
  $E_{37}Mark : p_3has\ note^{(p_{3.1}has\ type:E_{55}Type)} : E_{62}String$.

  Notation:

  $[\![p_{65}[\![\langle p_{3.1}, p_3\rangle]\!]]\!] : E_{22} \rightarrow \square(E_{37} \times \square(E_{62} \times E_{55}))$

- OCH (Condition/Examination History) is a narrative description of the condition or examination history of the art work. It is associated with $E_{14}Condition\ Assessment$ events, it can be mapped to $E_{62}String$ .Its relationship with CR can be expressed as:

  $CR.OCH = E_{22}Man-Made\ Object : p_{34}assessed\ by :$
  $E_{14}Condition\ Assessment : p_3has\ note^{(p_{3.1}has\ type:E_{55}Type)} : E_{62}String$.

Notation:

$$[\![p_{34}[\![\langle p_{3.1}, p_3\rangle]\!]]\!]: E_{22} \rightarrow \square\left(E_{14} \times \square\left(E_{62} \times E_{55}\right)\right)$$

- OTH (Treatment/Conservation History) is a narrative description of the treatment or conservation history of the art work. Mapping of OTH is similar to the mapping of OCH, refer to OCH for detailed mapping.

  Notation:

$$[\![p_{31}[\![\langle p_{3.1}, p_3\rangle]\!]]\!]: E_{22} \rightarrow \square\left(E_{11} \times \square\left(E_{62} \times E_{55}\right)\right)$$

In the following sections, "who made it", "when was it made", "where was it made" all appear in the CRM in one or more $E_{12}Production$, so all these fields are related $E_{12}Production$.

5.3.3.5 *Who made it?*

CRG (Creator Group) is the attribute used to group the fields associated with the creator of the art work. Doerr (2001a) suggests it maps to $E_{21}Person$, however, he also mentions that CRG can also map to $E_{39}Actor$, even though those do not necessarily have birth and death. The relationship between CRG and CR becomes:
$CR.CRG = E_{22}Man-Made\ Object : P_{31}was\ produced$ by:
$E_{12}Production : P_{14}carried\ out\ by : E_{39}Actor$

Notation:

$[\![p_{31}[\![p_{14}]\!]]\!]: E_{22} \rightarrow \square\left(E_{12} \times \square E_{39}\right)$.

The following fields are those subfields covered by CRG:

- CRQ (Creator-Qualifier) is the term used to describe the qualification of the attribution of the work to a particular creator. It is treated as the description of the person.
- CRT (Creator-Name-Text) displays the name of the creator.
- CDT (Creator-Dates/Locations-Text) is a description of the date and place associated with the creator of the art work.
- CRB (Creator-Biography) is a biographical description of the creator of the art work.
- CNO (Creator-Notes) is a text note about the creator, as well as the relationship between creator and the art work.

CRQ, CRT, CRN CDT CRB and CNO all maps to $E_{62}String$ in CRM.

- CRN (Creator-Name) displays in short form the creator's name. It maps to $E_{21}Person$ itself.
- CRC (Creator-Culture/Nationality) is the culture or nationality of the creator of the artwork. According to Doerr (2001a), three interpretations can be applied to CRC: the culture the creator was born into, this can be mapped to $E_{55}Type$ of Actor/Person or as the $E_4Period$ of creator's birth falls into; or CRC can be treated as the group he/she belonged to during the creation, this can be mapped to $E_{74}Group$; in the last situation, the CRC is the cultural context of the creation, it can be mapped as the period of the production falls into.
- CGN (Creator-Gender) represents the gender of a person, it maps to $E_{55}Type$.

The following fields (CBD, CBQ and CBP) are all related with the event $E_{67}Birth$:

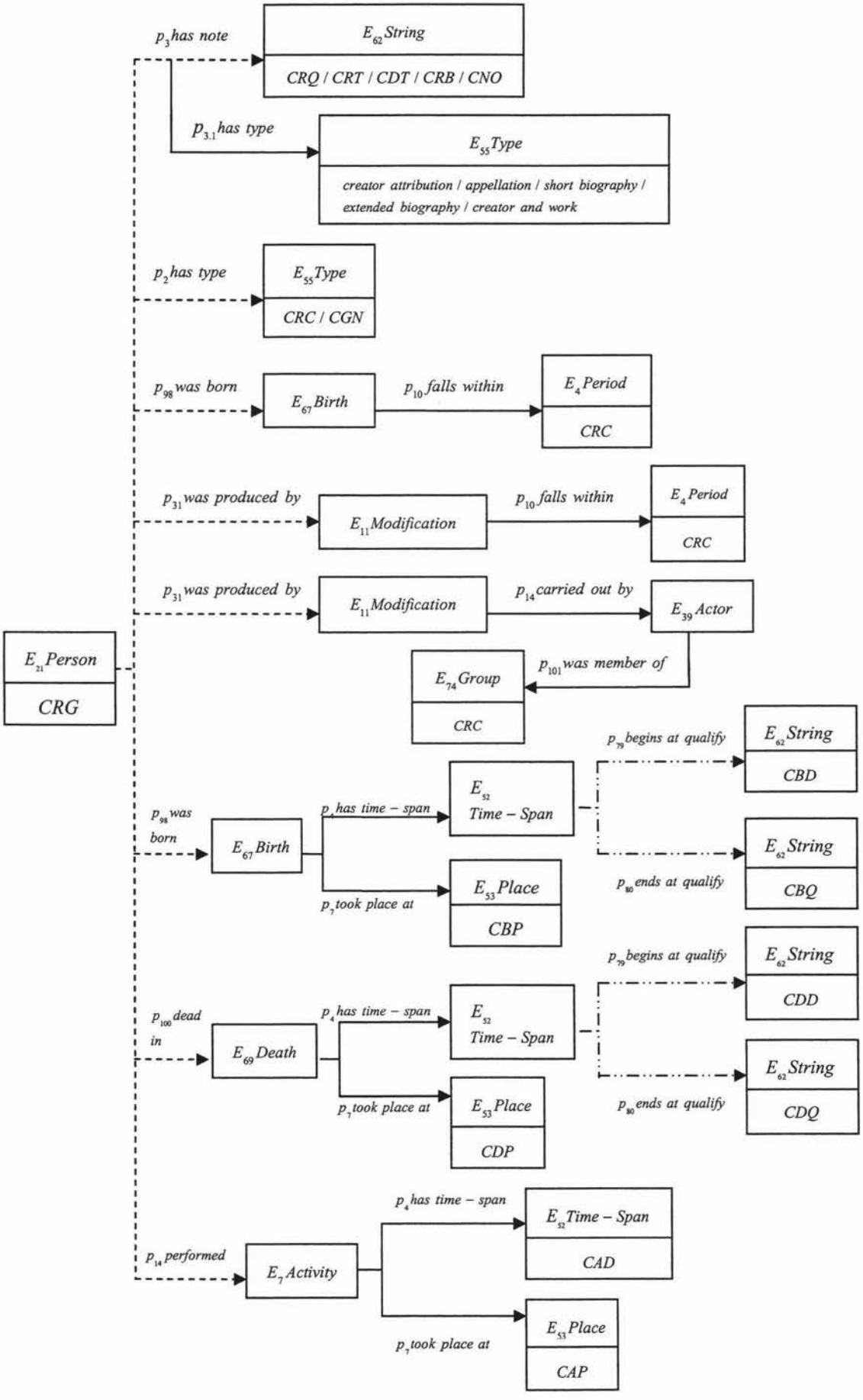
- CBD (Creator-Birth-Date) represents the date of the birth of the creator, it is a link of the $E_{67}Birth$ event. It maps to $E_{52}Time-Span$.
- CBQ (Creator-Birth-Qualifier) is a text qualifier for the creator's date of birth, such as the qualifiers of "before", "after" that used to limit the date. It can be mapped as $E_{62}String$.
- CBP (Creator-Birth-Place) represents the birth place of the creator. It maps as $E_{53}Place$.

The following fields (CDD, CDQ and CDP) are all related to the event $E_{69}Death$:

- CDD (Creator-Death-Date) represent the death date of the creator, similar with CBD, it maps to $E_{52}Time-Span$.
- CDQ (Creator-Death-Qualifier). Same with the CBQ, is a text qualifier and it maps to $E_{62}String$.
- CDP (Creator-Death-Place), is similar to the mapping of CBP. It also maps as $E_{53}Place$.
- CAD (Creator-Active-Date) is the creator's date of activity; it may correspondent to multiple activities. It is similar to the method of mapping the generic element "date".
- CAP (Creator-Active-Place) is the place where the creator is active; it may correspondent to multiple activities. It maps to $E_{53}Place$.

The CRG (Creator Group) is displayed in the following diagram:

Mapping diagram:



Notation:

$$[\![\langle p_3, p_{3.1}\rangle^5, p_2, p_{98}\langle p_{10}\rangle, p_{31}\langle p_{10}\rangle, p_{31}\langle p_{14}\rangle\langle p_{101}\rangle, p_{98}\langle p_7, p_4[p_{79}, p_{80}],\rangle, p_{100}\langle p_7, p_4[p_{79}, p_{80}]\rangle, p_{14}\langle p_4, p_7\rangle]\!] : E_{21} \rightarrow$$

$$\square\begin{pmatrix}(E_{62}\times E_{55})^5, E_{55}, E_{67}\times E_4, E_{11}\times E_4, E_{11}\times E_{39}\times E_{74}, E_{67}\times(E_{53}\times E_{52}\times(E_{62}\oplus E_{62})),\\ E_{69}\times(E_{53}\times E_{52}\times(E_{62}\oplus E_{62})), E_7\times E_{52}\times E_{53}\end{pmatrix}$$

(Note: at the beginning of the notation, the power of 5 has been assigned to $\langle p_3, p_{3.1}\rangle$ and $(E_{62}\times E_{55})$ as: $\langle p_3, p_{3.1}\rangle^5$ and $(E_{62}\times E_{55})^5$, as there are five AMICO fields mapped to $E_{62}String$, and they share the same notation as:

$[\![\langle p_3, p_{111}\rangle]\!] : E_{22} \rightarrow \square(E_{62}\times E_{55})$. This format will apply to rest of the notation under the same circumstances)

5.3.3.6 *When was it made?*

OCG (Creation-Dates) is the collections of fields represent in the date of the creation, OCG is corresponds to $E_{52}Time-Span$ relate to the $E_{12}Production$ event, the mapping of OCG is: $\begin{matrix}CR.OCG = E_{22}Man-Made\ Object : p_{31}was\ produced\ by :\\ E_{12}Production : p_4has\ time-span : E_{52}Time-Span\end{matrix}$.

Notation:

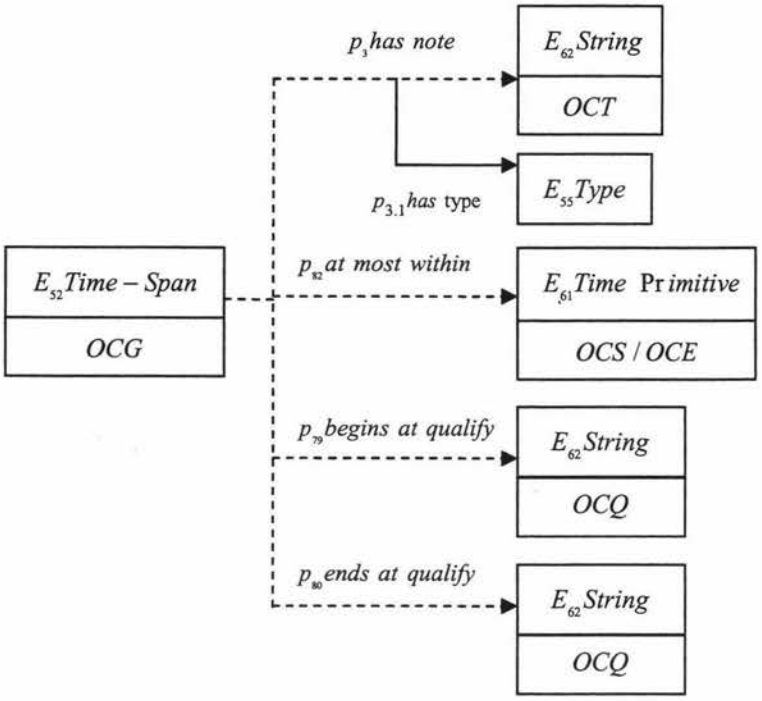
$[\![\langle p_{31}, p_4\rangle]\!] : E_{22} \rightarrow \square(E_{12}\times E_{52})$.

OCG covers fields OCT, OCS, OCE and OCQ.

- OCT (Creation-Date-Text) represents the related text on date when the work was created. It can be mapped as $E_{62}String$;
- OCS (Creation-Date-Start) and OCE (Creation-Date-End) both map to $E_{61}Time\ Primitive$.
- OCQ (Creation-Date-Qualifier) is a qualifier that indicates an approximation to the earliest or latest date. It maps to $E_{62}String$ in CRM.

The mapping of OCG and its subgroup here are similar to the mapping of the generic element "date", see the following diagram for mapping:

Mapping diagram:



Notation:

$$[\![\langle p_3, p_{3.1}\rangle, (p_{82})^2, p_{79}, p_{80}]\!] : E_{52} \rightarrow \square\left(\langle E_{62} \times E_{55}\rangle, (E_{61})^2, E_{62}, E_{62}\right).$$

5.3.3.7 Where was it made?

OCP (Creation-Place) represents the place (places) where the work was created, it corresponds to $E_{53}Place$, and it relates to the $E_{12}Production$ event. The relationship between OCP and CR is:

$CR.OCG = E_{22}Man-Made\ Object : p_{31}was\ produced\ by :$
$E_{12}Production : p_7 took\ place\ at : E_{53}Place$

Notation: $[\![\langle p_{31}, p_7\rangle]\!] : E_{22} \rightarrow \square(E_{12} \times E_{53})$

5.3.3.8 What is it about?

- STG (Style/Period Group) is the group contains the fields describing the style and period of the art work.

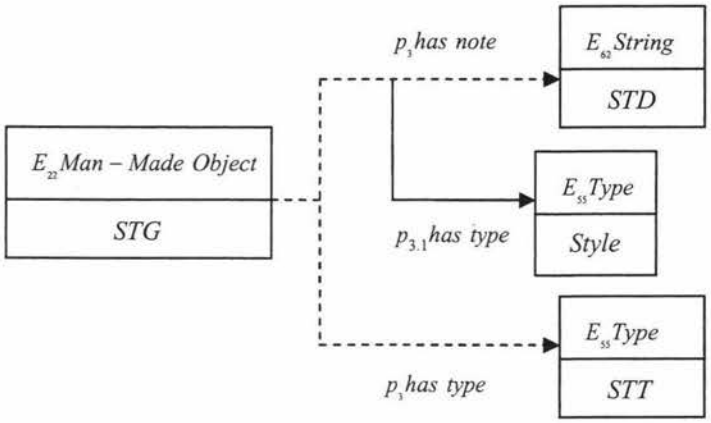
  STG covers STD (Style/Period-Description) and STT (Style/Period-Terms).

  - STD is a narrative description of the style or period of the art work, it maps to $E_{62}String$;

- STT is the index terms that represent the style or period of the work, it can be mapped as $E_{55}Type$.

Doerr (2001a) points out that style and period are not explicitly expressed in the CRM. The diagram below shows possible mapping of STG use $E_{62}String$ and $E_{55}Type$.

Mapping diagram:



Notation:

$[\![\langle p_3, p_{3.1}\rangle, p_2]\!] : E_{22} \rightarrow \square(E_{62} \times E_{55}, E_{55})$.

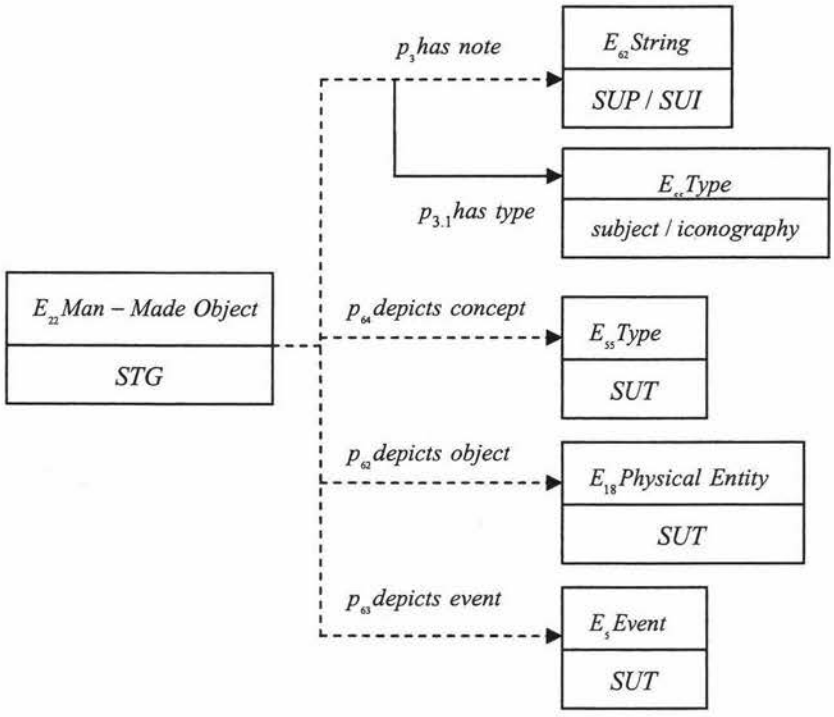
- SUG (Subject Matter Group) contains the fields documenting the work's subject matter.

  SUG contains fields: SUP (Subject Matter-PreIconographic Description), SUI (Subject Matter-Iconography) and SUT (Subject Matter-Index Terms).

  - SUP is a description of generic subject of the art work, it maps to $E_{62}String$;
  - SUI is a description of the specific, named subject of the art work, it maps to $E_{62}String$;
  - SUT is the index term that represent the subject of the art work. SUT maps to any combination of three different links depending on the kind of subject: $E_{55}Type$ (when SUT refers to non-real subjects), $E_{18}Physical\ Entity$ (when SUT refers to any living or dead object) and E_5Event (when SUT refers to event).

  SUG is "depicts" links and textual notes of the object, it doesn't map to dedicated entities in the CRM, using instead CRM $E_{62}String$ and $E_{55}Type$. See the mapping diagram below:

Mapping diagram:

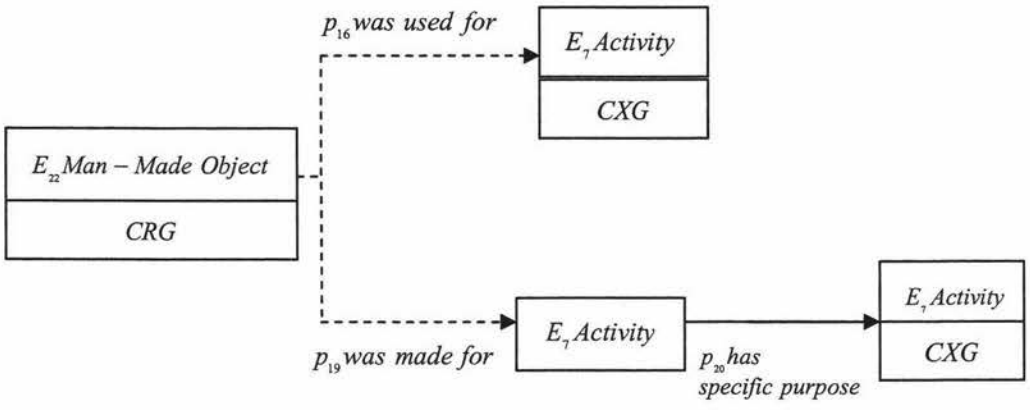


Notation:

$$[\![\langle p_3, p_{3.1}\rangle^2, p_{64}, p_{62}, p_{63}]\!] : E_{22} \rightarrow \square\left(\left(E_{62} \times E_{55}\right)^2, E_{55}, E_{18}, E_5\right)$$

- CXG (Context Group) corresponds to a series of E_7 *Actvity*, or it can be mapped as a text attached to the object. Mapping of CXG is shown in the following diagram:

  Mapping diagram:



Notation:

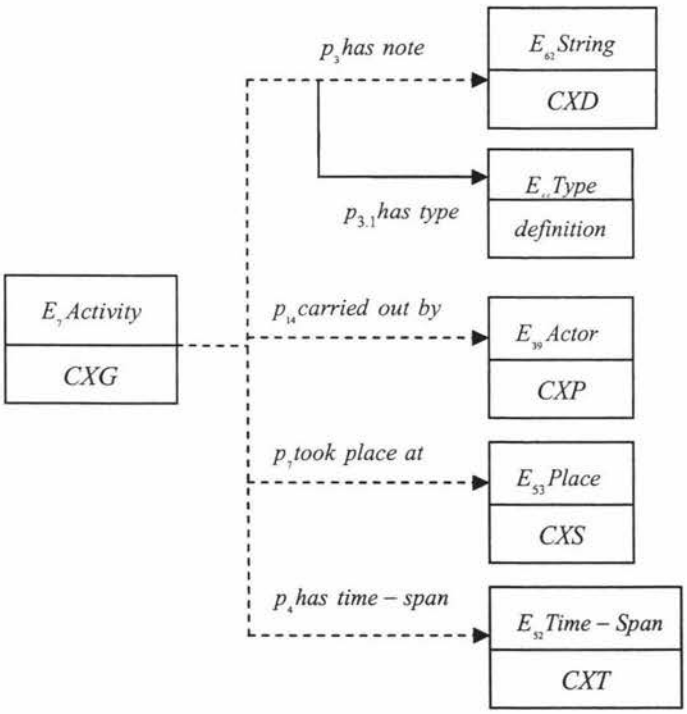
$$[\![\langle p_{19}, p_{20}\rangle, p_{16}]\!] : E_{22} \rightarrow \square\{E_7\}$$

CXG contains the fields documenting work's context, these fields are: CXD (Context-Description), CXP (Context-Related-Person), CXS (Context-Related Site/Place) and CXT (Context-Time Period/Dates).

- CXD is a description of the historic context of the art work, includes the art work's creation, display or other historical information. CXD can map to $E_{62}String$ as a text attached to the object itself, or the text attached to the individual activities.
- CXP is the index form of the name of any people related to the art work, it maps to $E_{39}Actor$.
- CXS represents names of any places that related to the art work, it can be mapped as $E_{53}Place$;
- CXT records the date, time or periods of a particular context, it can be mapped as $E_{52}Time-Span$.

See the diagram below for the detailed mapping for CXG:

Mapping diagram:



Notation:

$[\![\langle p_3, p_{3.1}\rangle, p_{14}, p_7, p_4]\!] : E_7 \rightarrow \square(E_{62} \times E_{55}, E_{39}, E_{53}, E_{52})$.

CXD can also be mapped as $E_{62}String$ attached to the object itself:

$E_{22}Man-Made\ Object : p_3 has\ note^{[has\ type:E_{55}Type(definition)]} : E_{62}String$.

Notation:

$[\![\langle p_3, p_{3.1}\rangle]\!] : E_{22} \rightarrow \square(E_{55} \times E_{62})$.

5.3.3.9 What does it mean?

OCR (Critical Responses) is a critical discuss about the art work. It can be mapped as $E_{62}String$:

$CR.OCR = E_{22}Man-Made\ Object : p_3 has\ note^{[has\ type:E_{55}Type]} : E_{62}String$

Notation:

$[\![\langle p_3, p_{111}\rangle]\!] : E_{22} \rightarrow \square(E_{55} \times E_{62})$.

5.3.3.10 Who showed it?

OEH (Exhibition or Loan History) is the record of history of when and where the work has been exhibited. It can be mapped as $E_7 Activity$ in CRM: $E_{22}Man-Made\ Object : p_{16} was\ used\ for : E_7 Activity : p_3 has\ type : E_{55}Type$.

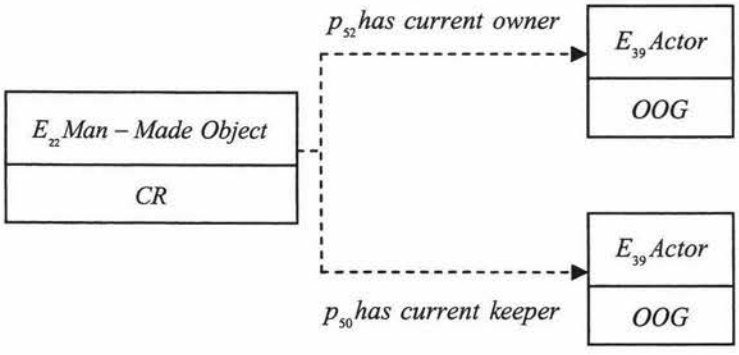
Notation:

$[\![p_{16}\langle p_3\rangle]\!] : E_{22} \rightarrow \square(E_7 \times E_{55})$.

5.3.3.11 Who owned it?

- OOG (Owner Group) used to group the fields documenting the ownership of the work. OOG corresponds to $E_{39}Actor$ in CRM, its relationship with CR shown in the following diagram:

  Mapping diagram:



  Notation:

  $[\![p_{50}, p_{52}]\!] : E_{22} \rightarrow \square(E_{39}, E_{39})$.

  OOG group contains fields: OON (Owner Name), OOP (Owner-Place), OOA (Owner-Accession-Number) and OOC (Owner-Credit-Line).

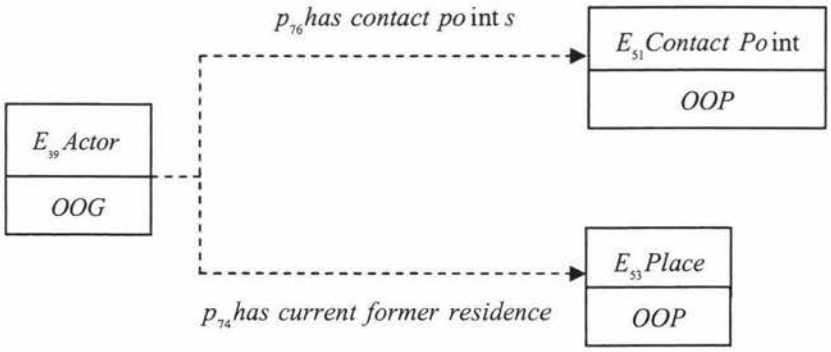
  - OON represents name of the owner, it maps to $E_{39}Actor$ itself.

- OOP is the place of ownership. It either maps to $E_{51}Contact\ Point$ attached to the owner or maps to $E_{53}Place$ as the link of actor's resides.
- OOA represents the accession number applied to the art work by the owner;
- OOC is any acknowledgement related to the ownership of the art work.

Both OOA and OOC are the fields associated with the object, so these two fields are related to $E_{22}Man-Made\ Object$ instead of with $E_{39}Actor$.

The mapping of OOG and its fields are demonstrated in the following diagram:

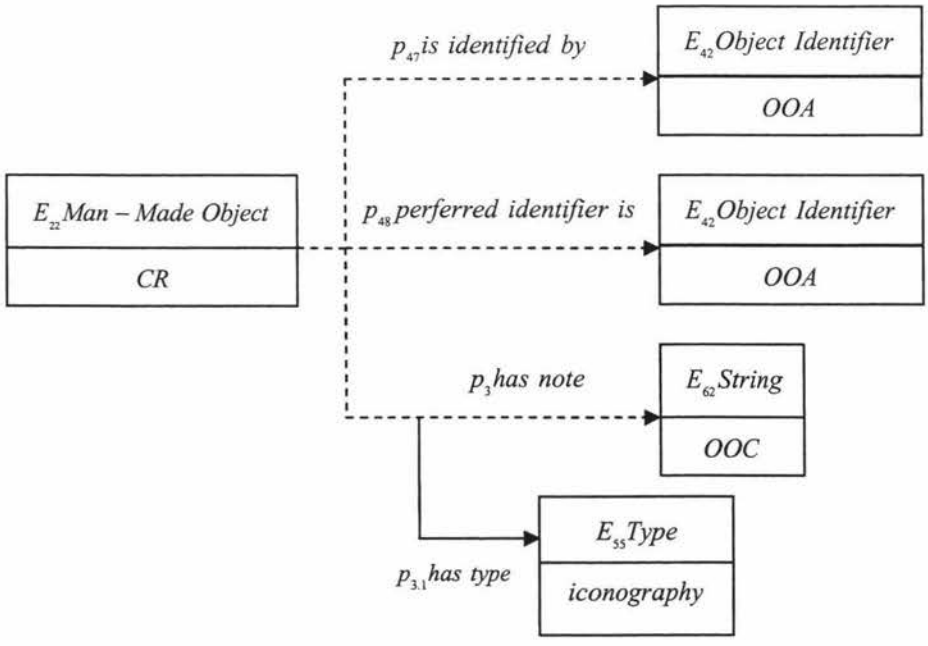
Mapping diagram:



Notation:

$[\![p_{76}, p_{74}]\!]: E_{39} \rightarrow \square(E_{51}, E_{53})$.

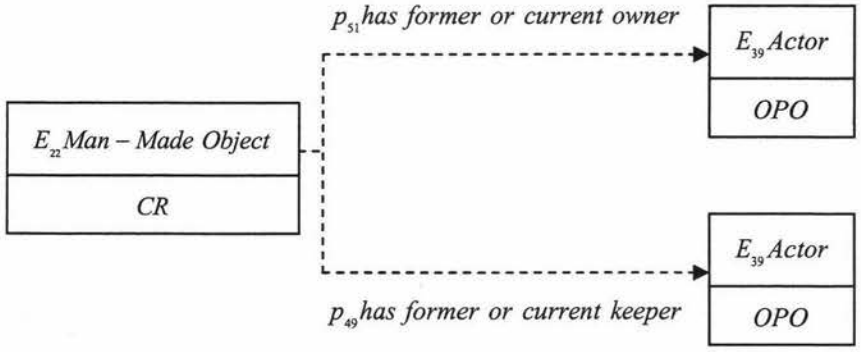
The mapping of OOA and OOC are associated with $E_{22}Man-Made\ Object$:



Notation:

$[\![p_{47}, p_{48}, \langle p_3, p_{3.1}\rangle]\!]: E_{22} \rightarrow \square(E_{42}, E_{42}, E_{62} \times E_{55})$

- OPO (Provenance/Prior Owners-Text) is a record of the previous owner of the art work. It maps to $E_{39}Actor$, the mapping can be described using the following diagram:



Notation:

$[\![p_{51}, p_{49}]\!]: E_{22} \rightarrow \square(E_{39}, E_{39})$

- ORG (Rights/Copyright) is a group used to represent the work's copyright or restriction, it maps to $E_{30}Right$ in CRM:

  $E_{22}Man-Made\ Object: p_{104}is\ subject\ to: E_{30}Right$.

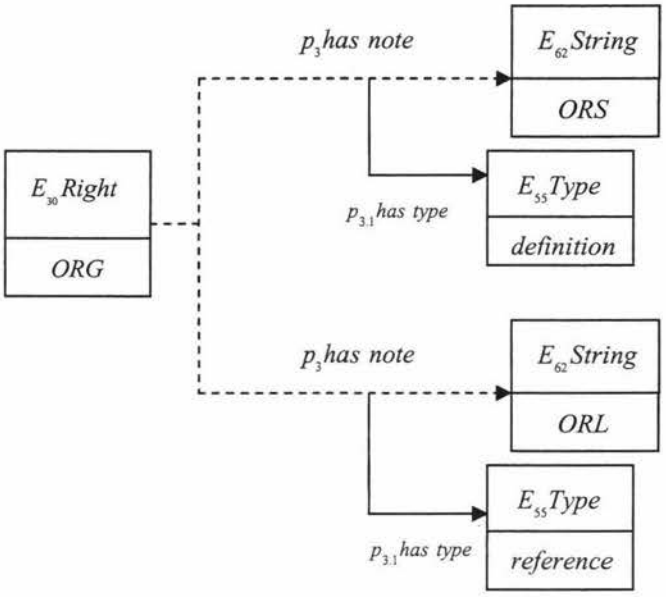
  Notation:

  $[\![p_{104}]\!]: E_{22} \rightarrow \square E_{30}$.

  The fields belong to ORG are: ORS (Copyright-Statement) and ORL (Copyright-Link).

  - ORS is the statement of the copyright, including any known copyright holders or restrictions;
  - ORL is the link to AMICO stuff to indicate how to get the further information of the copyright for the art work. Both ORS and ORL map to $E_{62}String$.

Mapping diagram:



Notation:

$$[\![\langle p_3, p_{3.1}\rangle^2]\!] : E_{30} \rightarrow \square(E_{62} \times E_{55})^2 .$$

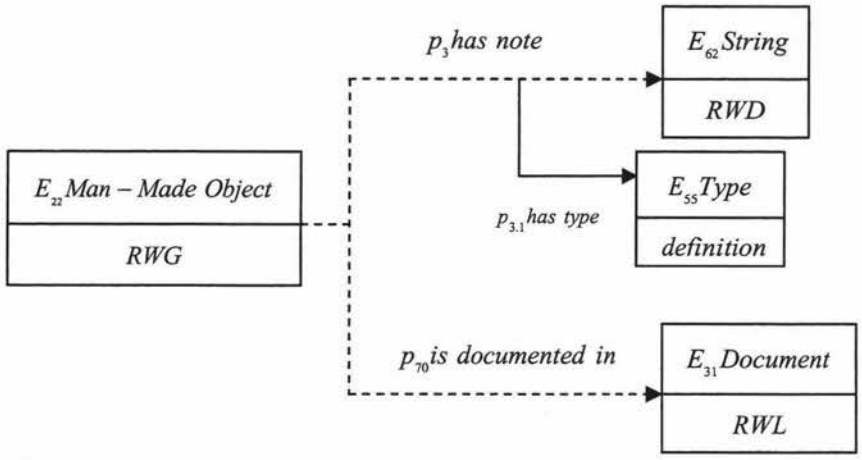
5.3.3.12 What is it related to?

- RWG (Related Works of Art) group fields documenting related works of art, it corresponds to $E_{22}Man-Made\ Object$ in CRM.

  The fields belong to RWG are: RWD (Related-Works-Description), RWL (Related-Works-Identifier/Link) and RWR (Related-Works-Relationship-Type).

  - RWD is a description of the relationship between this art work and others.
  - RWL is an identifier to the related art work.
  - RWR is a relationship drawn from the DC but it has an equivalent entitiy in the CRM, however, Doerr (2001a) suggests that RWR can be mapped to the Dublin Core relationship.

Mapping diagram:



Notation:

$[\![\langle p_3, p_{3.1}\rangle, p_{70}]\!] : E_{22} \rightarrow \square(\langle E_{62} \times E_{55}\rangle, E_{31})$

- RIG (Related Images Group) corresponds to $E_{38}Image$. The relationship between CR and RIG is:

  $E_{22}Man-Made\ Object : p_{67}is\ referred\ to\ by :$
  $E_{28}Conceptual\ Object : p_{67}refers\ to : E_{38}Image$

  Notation:

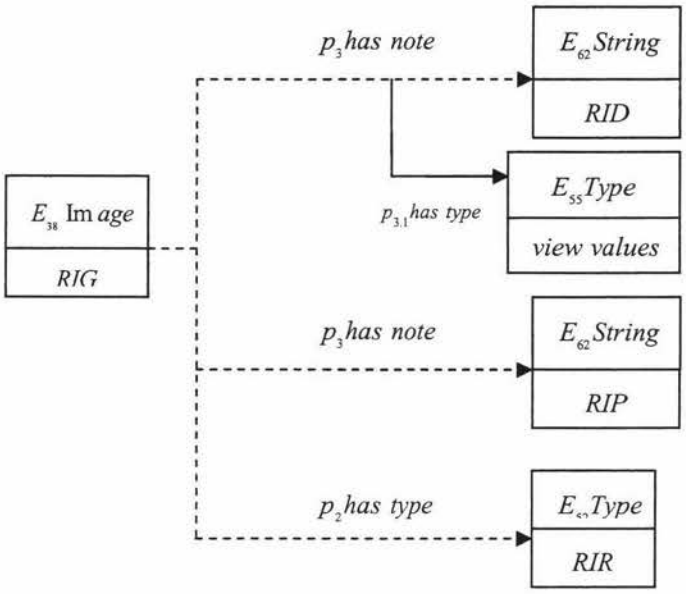
  $[\![p_{67}[\![p_{67}]\!]]\!] : E_{22} \rightarrow \square(E_{28} \times \square E_{38})$.

  All AMICO works must have at least one related image. RIG contains fields documenting related images, these fields are: RIL (Related-Image-Identifier/Link), RID (Related-Image-Description), RIP (Related-Image-Preferred) and RIR (Related-Image-Relationship-Type).

  - RIL is the identifier to the related images, it maps to $E_{38}Image$. As RIL is a multimedia data, according to AMICO, it also maps to the DC.Resource.Identifier in the Dublin Core. Refer to the mapping of DC.Resource.Identifier for more detailed mapping.
  - RID is the view of the work shown in the image, such as "Aerial View", "Full View"...etc. RID maps to $E_{62}String$. Because RID is a multimedia related record, it also maps to DC.Description in Dublin Core. Refer to the mapping of DC.Description for more detailed mapping;
  - RIP indicate "yes/no" whether it is the preferred image of the work, it maps to $E_{62}String$;
  - RIR represent the version/format of the image, it maps to $E_{55}Type$.

The mapping of RIG and its fields is shown in the following diagram:

Mapping diagram:



Notation:

$[\![\langle p_3, p_{3.1}\rangle, p_3, p_2]\!]: E_{38} \rightarrow \square(E_{62} \times E_{55}, E_{62}, E_{55})$.

- RMG (Related Multimedia) and RDG (Related Documents) are groups containing documenting related files. RMG is concerned with multimedia files; RDG covers both non-multimedia and multimedia files. Both RMG and RDG map to $E_{31}Document$, they have same mapping notation: $[\![p_{70}]\!]: E_{22} \rightarrow \square E_{31}$. RMG contains fields RML (Related-Multimedia-Identifier/Link), RMR (Related-Multimedia-Relationship-Type) and RMD (Related-Multimedia-Description).
  - RML is the identifier of the related multimedia file, it maps to $E_{31}Document$ itself. RML also maps to DC.Resource.Identifier, refer to mapping of DC.Resource.Identifier in Dublin Core for details.
  - RMR represent the relationship between the art work and related multimedia files. RMR does not map to CRM. AMICO recommends that RMR map to DC.ResourceType in Dublin Core, refer to mapping of DC.ResourceType for more detailed mapping.
  - RMD is a description of the related multimedia files, it maps to $E_{62}String$ in CRM, and the mapping notation is: $[\![p_2]\!]: E_{31} \rightarrow \square E_{62}$. RMD also maps to DC.Description, refer to mapping of DC.Description in Dublin Core for details.

- RDG (Related Documents) contains fields RDL (Related-Document-Identifier/Link), RDD (Related-Document-Description) and RDR (Related-Document-Relationship-Type).
- RDL is the identifier of the related file, it maps to $E_{31}Document$ itself. RDL maps to DC.Resource.Identifier when RDL represents multimedia metadata record, refer to mapping of DC.Resource.Identifier in Dublin Core for details.
- RDR represent the relationship between the art work and related document, similar to RMR. RDR does not map to CRM, AMICO recommends that RMR map to DC.ResourceType in the Dublin Core, refer to mapping of DC.ResourceType for more detailed mapping.
- RDD is a description of the related document, it maps to $E_{62}String$ in CRM, and the mapping notation is: $[\![p_2]\!]: E_{31} \rightarrow \square E_{62}$. RDD also maps to DC.Description when RML represents refers to multimedia meta record, refer to mapping of DC.Description in Dublin Core for details.

5.3.3.13 Who documented it?

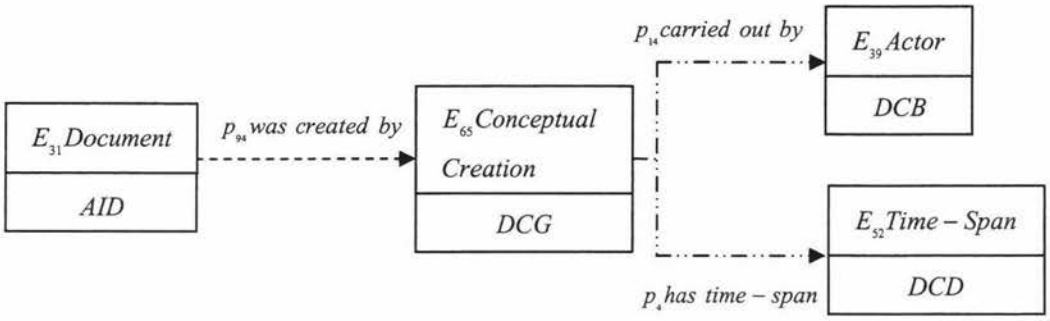
- DCG (Documentation/Cataloguing-History Group) is a group that contains fields recording the documentation history of the work.

  DCG contains fields: DCB (Documented/Cataloged By) and DCD (Documented/Cataloged-Date). As AID maps to the AMICO catalogue record, DCG maps to $E_{65}Conceptual\ Creation$ event that is associated with the AID document.

  - DCB is the name of the person who documents the work, it maps to $E_{39}Actor$.
  - DCD records the date the work was documented, it maps to $E_{52}Time-Span$.

  The mapping of DCG and its subgroup is shown in the following diagram:

Mapping diagram:



Notation:

$$[\![p_{94}[p_{14}, p_4]]\!] : E_{31} \rightarrow \square\left(E_{65} \times \left(E_{39} \oplus E_{52}\right)\right)$$

5.4 Summary

The use of the mapping diagram and category theory notation to replicate the mapping of EAD, Dublin Core and AMICO was successful. All mapping situations were able to be expressed. Based on the success of these two it is conducted that the research aims 2 and 3 were achieved, namely "develop tools to in the validation of the CRM" and "replicate the validation of the CRM on an international level". Mapping of EAD was based on the understanding and anlysis of the described archived material; mapping of Dublin Core was undertaken by transforming the unspecific metadata standard to the more precise CRM; mapping of AMICO was based on the analysis of an AMICO record and its categoires. These mappings demonstrate the power of the CRM approach and the interoperability of the CRM. The mapping is a successful proof of the CRM concept and shows that it is possible to preserve the meaning of the original information. However, the more analytic is the data record, the deeper the reasoning mapping is able to reach. Doerr (2001a) suggests that it may be worthwhile to apply a standardised mapping note-type for interoperability of the various text types; this suggestion made by Doerr, was an incentive to develop the mathematical notation used in the chapter, which can be considered as a first step towards making a standardised mapping note-type.

6 Analysis 3: New Zealand Validation of the CRM (Two New Zealand Studies)

6.1 Introduction

This chapter addresses the fourth aim of the research, which is to apply the same graphical and mathematical notation to validate the CRM within the New Zealand context. Two mapping activities have been applied to two organisations of New Zealand. The activities aim to improve the theoretical understanding of CRM, and validate the model into practice. The mapping practices, suggested by Doerr (2001b), seek to validate the CRM against local and international collections. The two New Zealand organisations, which are the subject of this validation process are: Suter Gallery in Nelson (South Island) and the Te Manawa Museum in Palmerston North. To validate the interoperability of CRM these two organisations were selected by the researcher as they involve different collection types and their data cataloguing systems are built under totally different database structures.

6.2 Mapping of Suter Gallery Data Model to CRM Version 3.0

6.2.1 Introduction of Suter Gallery collection and Data Model

The Suter Gallery has the third oldest collection in the South Island of New Zealand. Works of national significance include works by nineteenth century watercolour artist John Gully, works by Sir Tosswill Woollaston (one of the founders of modern art in New Zealand), and ceramics by local and national artists (refer to Appendix 4).

In comparison with modern standard, the Suter collections management system is seen as basic. Sorting on fields is the key facility. The record-based system was designed and developed by Suter.

The fact that Suter Gallery is a medium-size gallery and has had little exposure to e-technology, there is no internet access to the collections. In order to produce a better mapping result, the researcher divided Suter data fields into different categories based on the meaning the fields represented. The researcher added some new fields and grouped them into new categories, such as 'Copyright Group'. This group included fields such as 'Obj Constraint' and 'Obj Copyright'.

The data has some similarity to that of AMICO, this is reflected in the choice of category. However, while there are numbers of image files in the Suter Gallery; not every work has its own image. In this mapping, the fields for those image files are mapping to the corresponding entities in Dublin Core – a process adopted in AMICO.

A fully described data record field is shown in Table 4 below, the description of each of the data fields is based on Suter's file "Collection Database Fields".

Table 4 Suter Gallery Data Fields

Suter Fields	Group	Definition
Unique Identification		
SR (Suter Record)		Refers to the instance of the Suter object.
Number		Auto number, assigned to a work in the Suter record as a unique identifier
Accession No.		Number from the original Accession register.
		This must be a number. No letters or other characters may be used.
What is it?		
Object		Describes what object is: Sketch, Drawing, Ceramic etc.
No. of Pieces		Indicates the number of pieces in the complete work.
What is it called?		
Work Title		Name of work as given by artist.
What does it look like?		
Materials and Techniques	Group	Used to group fields documenting materials and techniques used to create the work.
Glaze	Materials and Techniques	This field refers to ceramic works only. Describes what glaze is used and glaze colours.
Medium	Materials and Techniques	Indicates material used to produce work; Oil, Acrylic, Crayon etc.
Description	Materials and Techniques	Contains descriptive material as necessary.
Support	Materials and Techniques	Describes what the work was produced on; Canvas, Paper, Board etc. Use one indicator only i.e. the primary support.
Support Auxiliary	Materials and Techniques	Describes material paper or canvas may be attached to i.e. paper (glued on card) should be described by one term only in this case paper (in support) and then Card in additional support. Additional materials (eg Glue) should be part of construction
Construction	Materials and Techniques	Describes how the object is constructed.
Decoration		This field refers to ceramic works only. Type of decoration used if any.
Measurement	Group	Used to group fields recording measurements.
Dim Height (mm)	Measurement	Dimension Height in millimetres.
Dim Width (mm)	Measurement	Dimension Width in millimetres.
Dim Depth (mm)	Measurement	Dimension Depth in millimetres.
Dim Diameter (mm)	Measurement	Dimension Diameter in millimetres.
Obj Condition		Records general notes on condition of work.
Obj Conservation Report		Contains detailed notes on damage, drying, cracking etc.
Obj Treatment		Contains detailed notes on treatment methods required to restore work to original condition.
Inscription		Marks or signature put on work by artist. Also includes Potters marks on Pottery works
When was it made?		
Creation Dates	Group	Used to group fields dating work's creation.
Date Created	Creation Dates	Date work was finished.

Suter Fields	Group	Definition
Who owned it?		
Provenance		Describes how the works first came to be in the Suter Collection; Donated, Bequeathed etc.
Donor Name		Records original donors name.
Obj Sponsor		Records names and date of conservation sponsors
Copyrights	Group	Records any Copyright conditions, which may apply to reproduction of work.
Obj Copyright	Copyrights	It is the statement of the copyright of the work.
Acquisition	Group	Used to group the fields related with the acquisition of the work.
Obj Mode of Accession	Acquisition	Records method of Accession; A=Allocated, L=Loan, D=Donated, Presented or Gifted, B=Bequeathed, P=Purchased
Obj Date of Accession	Acquisition	Refers to the date when the gallery get the work.
Obj Deaccession	Acquisition	Refers to the purpose of transferring the work from the gallery.
Obj Deaccession Date	Acquisition	Records date work is deaccessioned or removed from collection.
Obj Disposal	Acquisition	Records method of disposal.
Not located	Acquisition	Item missing from collection, no note, date or explanation for deaccession available.
What is it about?		
Artwork Exhibitions		Records any exhibitions the work has been used in. Should give the name of the exhibition, date and place.
Obj Constraints		Records any constraints, which may apply to the exhibition of any work.
Who made it?		
Creator	Group	Used to group fields documenting the creator of the work.
Aritist	Creator	Contains artists name and initials. No punctuation to be used.
Courtesy Title	Creator	Sir, Lady etc.
Forenames	Creator	Forenames where known.
Honours	Creator	Records any honours, which may have been bestowed on artist.
Date of Birth	Creator	Records Year of Birth.
Date of Death	Creator	Records Year of Death.
Nationality	Creator	Records nationality or place of birth.
Biographical Details	Creator	This field contains a brief biography of the artist.
Chronology	Creator	Contains career info in chronological order and may stand in place of biographical details.
Publications	Creator	Refer to a text description of the article published by the creator.
Artist Exhibitions	Creator	Records any exhibitions the artist has had works in.
What is it related to?		
Related Document	Group	Used to group the fields documenting related work of art
References	Related Document	Contains reference to any published material about the work.
Notes	Related Document	This field is a catchall for any information, which does not readily fit into any other field.
Related Images	Group	Used to group the fields documenting related images of the work.
Photo References	Related Images	Contains photo references where available.
Picture	Related Images	Contains graphical reproduction of work for reference purposes.
Location		Records location.
Price	Group	Used to group the fields recording the price related to the object.
Purchase price	Price	Price paid at time of purchase
Valuation for insurance	Price	Valuation based on market valuation of similar works sold within the last 4 years.
Loans	Price	Records any loans of works. Includes dates, reasons, and places.
Who documened it?		
Documentation	Group	Used to group the fields recording the documentation history of the work.
Cataloguer/Data Entry	Documentation	Name of data entry person.
Date Catalogued	Documentation	Date of last entry.
Inventory	Documentation	Refers to the updated by physical sighting of a work and record the date the object was last seen on.
Media Metadata Fields		
Photo: Photographer		Name of Photographer
Photo: Medium/Index		A numbering system for photographs/negatives/transparencies used to identify the photo collection.

6.2.2 Mapping formalism

The Suter data mapping approach is based on the CRM Version 3.0. The mapping formalism adopted is that used in the previous three validation exercises. Refer to section 5.1.2.2 for details.

6.2.3 Mapping Suter Gallery Data to CRM

Similar to the mapping of AMICO data, mapping takes place at two semantic levels. The first level views the completed record as an object; it can be treated as an instance of $E_{31}Document$, which can be identified by 'Number' and its 'Accession Number' as the data identifier. The second level is the actual content of the Suter dataset about an object in terms of its data fields to which it refers and relates.

6.2.3.1 Mapping for Unique identification

- Each of the Suter catalogue records is in one to one correspondence with the described object. The Suter Catalogue Record (SCR) can be mapped to the CRM as follows:

  $SCR = E_{22}Man-Made\ Object$

- The Number field of the Suter dataset object can be mapped as an instance of $E_{31}Document$ that is used to identify the related object content:

  $Number.SCR = E_{31}Document : \overline{p_{70}}documents : E_{22}Man-Made\ Object$

  Notation:

  $\langle \overline{p_{57}} \rangle : E_{31} \rightarrow E_{22}$

- Accession Number is the number that is designated to the work itself. It can be mapped as $E_{42}Object\ Identifier$ to SCR ($E_{22}Man-Made\ Object$), here is the mapping:

  $Accession\ Number = E_{22}Man-Made\ Object : p_{47}is\ identified\ by : E_{42}Object\ Identifier$.

  Notation:

  $\langle p_{47} \rangle : E_{22} \rightarrow E_{42}$.

6.2.3.2 Mapping for "What is it?"

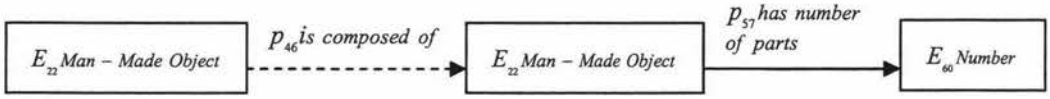
- Object corresponds to the subclass of $E_{55}Type$ for $E_{22}Man-Made\ Object$:

  $SCR.Object = E_{22}Man-Made\ Object : p_2has\ type : E_{55}Type$.

Notation:

$[\![p_2]\!]: E_{22} \rightarrow \square E_{55}$.

- Number of pieces represents the number of pieces of the object, it can be mapped as $E_{60}Number$. The correspondence relationship between SCR and 'Number of pieces' can be expressed using the following diagram:



Notation:

$[\![p_{46}\langle p_{57}\rangle]\!]: E_{22} \rightarrow \square\{E_{22} \times E_{60}\}$.

6.2.3.3 Mapping for "What is it called?"

Work Title maps as the title of the object, its correspondence relationship with SCR is: $SCR.Work\ Title = E_{22}Man-Made\ Object: p_{102}has\ title: E_{35}Title$.

Notation:

$[\![p_{102}]\!]: E_{22} \rightarrow E_{35}$

6.2.3.4 Mapping for "What does it look like?"

- Materials and Techniques Group (MTG) is used to present the group associated with the production process.

  The subfields of Materials and Techniques Group includes: Glaze, Medium, Description, Support, Support Auxiliary and Construction.

  MTG can be mapped as $E_{12}Production: MTG = E_{12}Production$, when it links with the Suter Catalogue Record, its relationship can be described as:

  $CR.OMG = E_{22}Man-Made\ Object: p_{31}was\ produced\ by: E_{12}Production$.

  Notation:

  $[\![p_{31}]\!]: E_{22} \rightarrow E_{12}$.

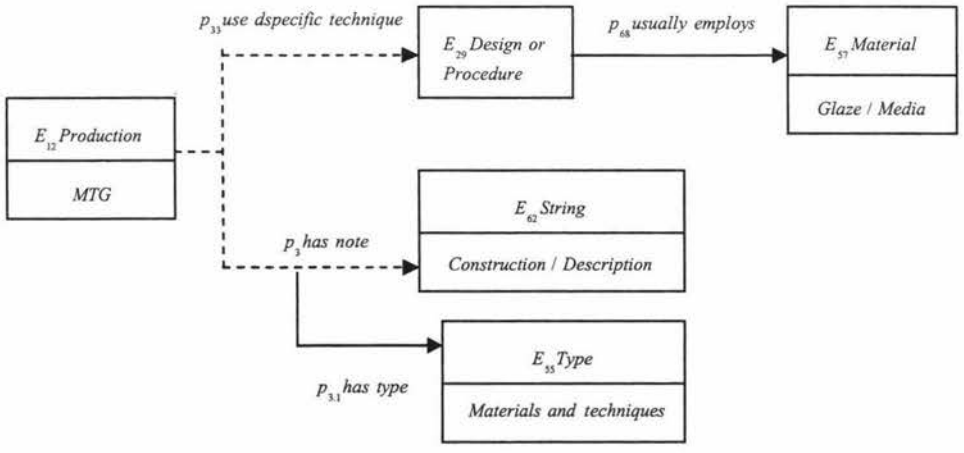
- Glaze is the material applied to the surface of the object. The reason for relating Glaze with Production here is because Glaze can be treated as the material used to create the object by applying specific technique, instead of associating glaze with the object. From the mapping of Glaze, it is found that professional knowledge of museum category is required; otherwise problems may occur when applying the mapping to the real situation. Glaze maps

to $E_{57}Material$, it is associated with either Suter Catalogue Record or Materials and Techniques Group.

- Medium is the term that indexes the materials used to create the work. Same with Glaze, it can be mapped as $E_{57}Material$, and it is associated with either Suter Catalogue Record or Materials and Techniques Group.
- Description is the term contains descriptive material as necessary. It maps to $E_{62}String$.
- Support is a single terms that indexes the support on which the work was created. It maps to $E_{57}Material$, and is associated with the Suter Catalogue Record.
- Support Auxiliary is the secondary support of the object; it can be treated as a free text description of the main Support. It maps to the subclass of $E_{55}Type$ for Support ($E_{57}Material$)
- Construction is related to how the work is made. It can be seen as a free text description of the materials and techniques used to create the work. So it can be mapped as $E_{62}String$ to the $E_{12}Production$.

The following diagram shows the mapping of MTG ($E_{12}Production$) and its related fields:

Mapping diagram:

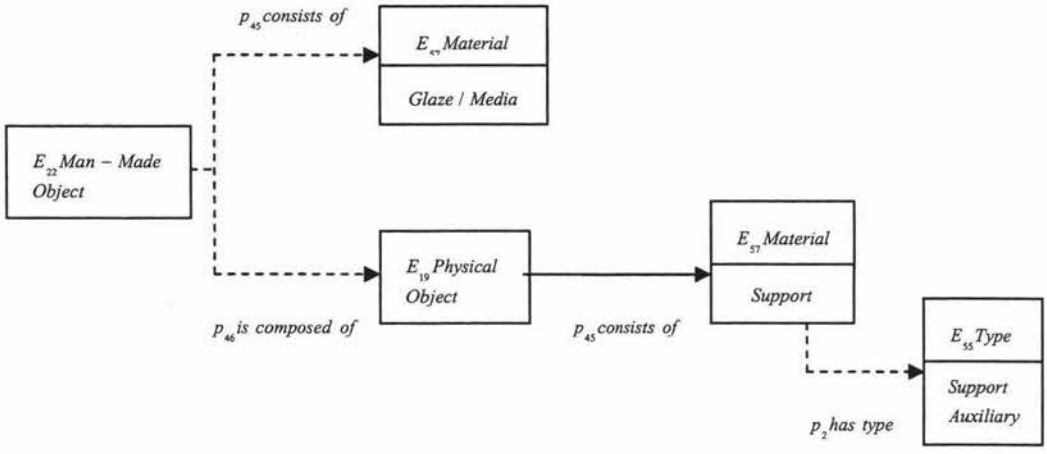


Notation:

$$[\![p_{33}\langle p_{68}\rangle^2, (p_3, p_{3.1})^2]\!] : E_{12} \rightarrow \square\left(E_{29} \times (E_{57})^2, (E_{62} \times E_{55})^2\right)$$

Fields related with MTG ($E_{12}Production$), such as 'Glaze', 'Support' and 'Support Auxiliary' all have the connection with Suter Catalogue Record ($E_{22}Man-Made\ Object$)

Mapping diagram:



Notation:

$$[\![(p_{45})^2, p_{46}[\![p_{45}[\![p_2]\!]]\!]]\!]: E_{22} \rightarrow \square((E_{57})^2, E_{19} \times (\langle E_{57} \rangle \times \square E_{55})).$$

- Decoration is a physical description of the object, it associates directly with the object, it can be mapped as $E_{62}String$:

  $SCR.Notation = E_{22}Man-Made\ Object : p_3 has\ note : E_{62}String$.

  Notation:

  $[\![p_3]\!]: E_{22} \rightarrow \square E_{62}$.

- Measurements Group is used to group the fields recording measurements of the object.

  Measurements Group include: Dim Height, Dim Width, Dim Depth and Dim Diameter.

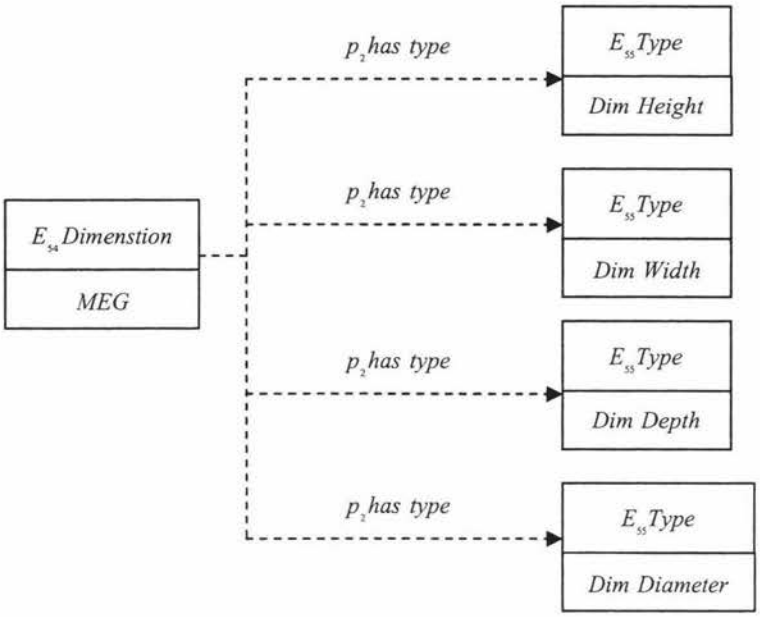
  Measurements Group maps to $E_{54}Dimension$. Its correspondent relationship with Suter Catalogue Record ($E_{22}Man-Made\ Object$) can be expressed as:

  $SCR.Measurement\ Group = E_{22}Man-Made\ Object : p_{43} has\ dimension : E_{54}Dimension$

  Notation:

  $[\![p_{43}]\!]: E_{22} \rightarrow \square E_{54}$

Dim Height, Dim Width, Dim Depth and Dim Diameter respectively correspond to the types measurement in the dimensions of Height, Width, Depth and Diameter; they all map to $E_{55}Type$ in CRM. The diagram below can be used to describe mapping of Measurements Group and its related fields:



Notation:

$[\![p_2, p_2, p_2, p_2]\!]: E_{54} \rightarrow \square(E_{55}, E_{55}, E_{55}, E_{55})$

- Obj Condition is a narrative description of the condition or examination history of the work. It associates with $E_{14}Condition\ Assessment$ events, it can be mapped to $E_{62}String$. Mapping of the Obj Condition is comparable with the mapping of OCH in AMICO, refer to OCH mapping for more details, the notation of mapping of Obj Condition can be expressed as:

$[\![p_{34}[\![\langle p_{3.1}, p_3\rangle]\!]]\!]: E_{22} \rightarrow \square(E_{14} \times \square(E_{62} \times E_{55}))$

- Obj Conservation Report is a description of how to look after and maintain the art work. It can be treated as the free text attached to the Object, and it can be mapped as $E_{62}String$:

$SCR.Obj\ Conservation\ Report = E_{22}Man-Made\ Object : p_3 has\ note : E_{62}String$

Notation:

$[\![p_3]\!]: E_{22} \rightarrow \square E_{62}$.

- Obj Treatment is a description of the methods used to repair damage to the work. It corresponds to a set of $E_{11}Modification$ events, every treatment is a

modification. It maps to $E_{62}String$ of the $E_{11}Modification$, refer to OTH in AMICO.

Notation:

$[\![p_{31},\langle p_{111},p_3\rangle]\!]: E_{22} \rightarrow \square\{E_{11}\times E_{62}\times E_{55}\}$

- Inscription is a free text description or transcription of any inscriptions or marks on the work. It corresponds to OIN in AMICO. Refer to the mapping of OIN for detailed Inscription mapping to CRM.

6.2.3.5 *Mapping for "When was it made?"*

Creation Dates Group is the collection of fields which represent the date of creation.

Creation Date Group contains fields: Date Created.

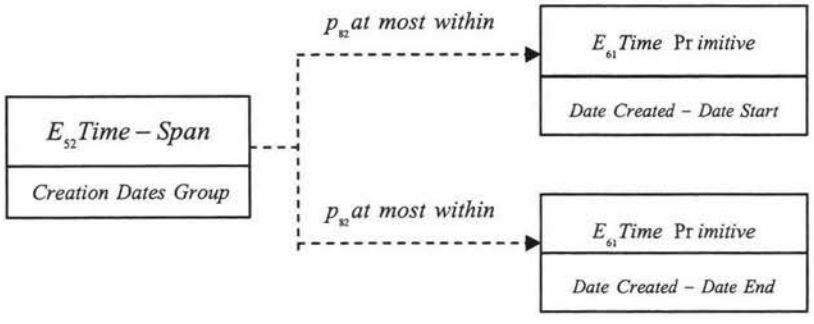
Creation Dates Group corresponds to $E_{52}Time-Span$ which is related to the $E_{12}Production$ event. Its relationship with the Suter Catalogue Record ($E_{22}Man-Made\ Object$) corresponds to the mapping of OCG in AMICO.

Notation:

$[\![\langle p_{31},p_4\rangle]\!]: E_{22} \rightarrow \square(E_{12}\times E_{52})$.

- Date Created indicates the time period over which the art work was created; it covers the Date-Start and Date-End of the creation of the work. It maps to $E_{61}Time\ Primitive$.

Diagram below shows the mapping of the Creation Dates Group and its fields:



Notation:

$[\![(p_{82})^2]\!]: E_{52} \rightarrow \square(E_{61})^2$

6.2.3.6 Mapping for "Who owned it?"

- Provenance refers to a record of the past owners of the art work. It corresponds to the OPO in AMICO, and is related to $E_{39}Actor$ (the previous owners). Refer to OPO mapping in AMICO for more detailed mapping.

  Notation:

  $[\![p_{50}, p_{52}]\!] : E_{22} \rightarrow \square(E_{39})$;

- Donor Name represents name of the donor. It maps to $E_{39}Actor$ itself, Donor Name mapping is same as Provenance.
- Object Sponsor is the name of the person who helps on sponsoring the work. Same with the mapping of Provenance and Donor. It maps to $E_{39}Actor$ itself.
- Copyrights Group is a group used to represent the work's copyright or restriction, it maps to $E_{30}Right$ in CRM, it has the same mapping with ORG in AMICO, its connection with Suter Catalogue Record ($E_{22}Man-Made\ Object$). Refer to mapping of ORG in AMICO for more details.

  Notation:

  $[\![p_{104}]\!] : E_{22} \rightarrow \square(E_{30})$.

  The fields belong to Copyright Group are: Obj Constraint and Obj Copyright.

  - Obj Copyright is the statement of the copyright. Both Obj Constraint and Obj Copyright map to $E_{62}String$.

  Refer to the mapping diagram of ORG in AMICO for detailed mapping.

  Notation for mapping of Copyright Group:

  $[\![\langle p_3, p_{3.1}\rangle]\!] : E_{30} \rightarrow \square(E_{62} \times E_{55})$

- Acquisition Group is the group used to indicate the beginning and the end of an ownership.

  The Acquisition Group covers the field of Obj Mode of Accession, Obj Date of Accession, Obj Deaccession, Obj Deaccession Date and Obj Disposal. Acquisition Group maps to $E_8Acquisition$ in CRM, its correspondent connection with Suter Catalogue Record ($E_{22}Man-Made\ Object$) is:

  $SCR.Acquisition\ Group =$

  $E_{22}Man-Made\ Object : p_{24}changed\ ownership\ by : E_8Acquisition$

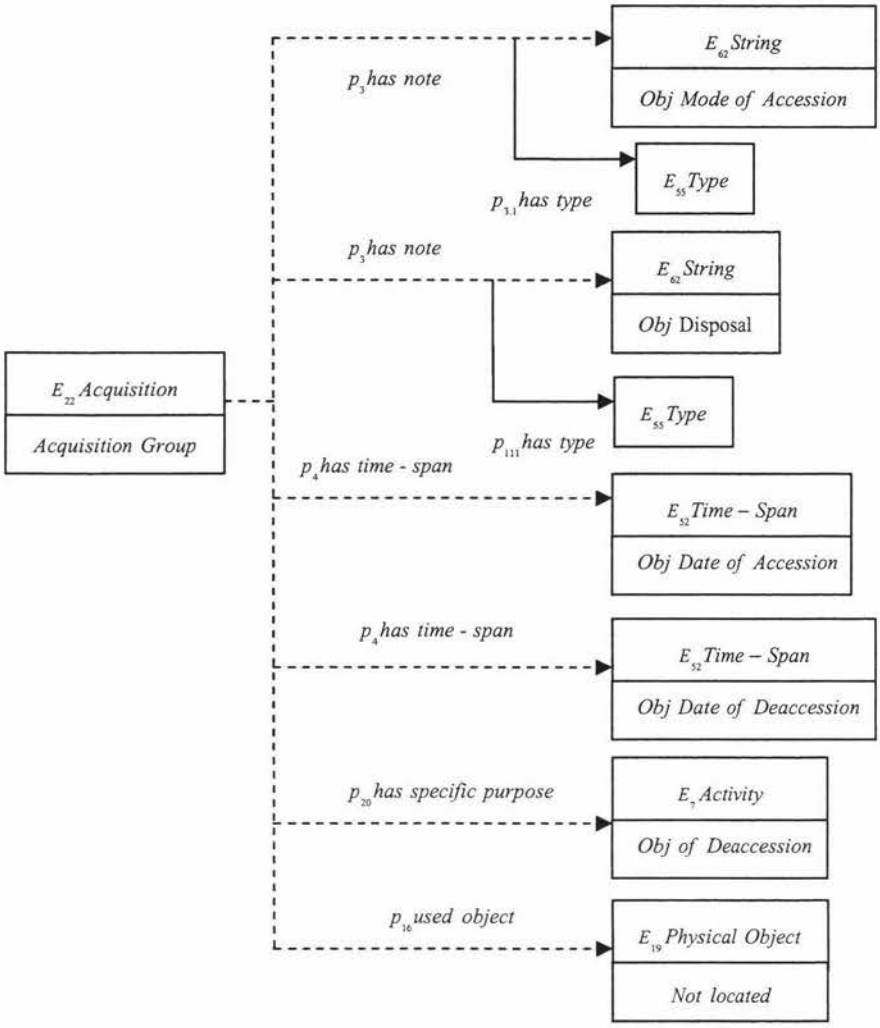
Notation:

$[\![p_{24}]\!]: E_{22} \rightarrow \square E_8$

- Obj Mode of Accession indicates in which way the work was from. It can be mapped as $E_{62}String$.
- Obj Date of Accession refers to the date when the gallery obtains the work. It can be mapped as $E_{52}Time-Span$.
- Obj Deaccession refers to the purpose of transferring the work from the gallery. It can be mapped as $E_7 Activity$.
- Obj Deaccession Date refers to the date when the work has been transfer to somewhere else from the gallery. It maps to $E_{52}Time-Span$ as well.
- Obj Disposal records method of disposal. It can be mapped as $E_{62}String$.
- Not located refers to the item missing from collection, no note, date or explanation for deaccession available. It maps to $E_{19}Physical\ Object$.

The following diagram shows the mapping of Acquisition Group and its related fields.

Mapping diagram:



Notation:

$$[\![\langle p_3, p_{3.1}\rangle^2, (p_4)^2, p_{20}, p_{16}]\!] : E_8 \rightarrow \square\left((E_{62} \times E_{55})^2, (E_{52})^2, E_7, E_{19}\right)$$

6.2.3.7 *Mapping for "What is it about?"*

- Artwork Exhibition is a narrative description of the historical context of the work of art, including its creation, display and exhibition. The mapping of Artwork Exhibition corresponds with the CXD mapping in AMICO, refer to CXD mapping for more details.

  Notation:

  $$[\![\langle p_3, p_{111}\rangle]\!] : E_{22} \rightarrow \square\{E_{62} \times E_{55}\}$$

- Obj Constraints refers to any constraints, which may apply to the exhibition of any work. It maps to $E_{62}String$.

Notation:

$[\![\langle p_3, p_{111} \rangle]\!] : E_{22} \rightarrow \square\{E_{62} \times E_{55}\}$

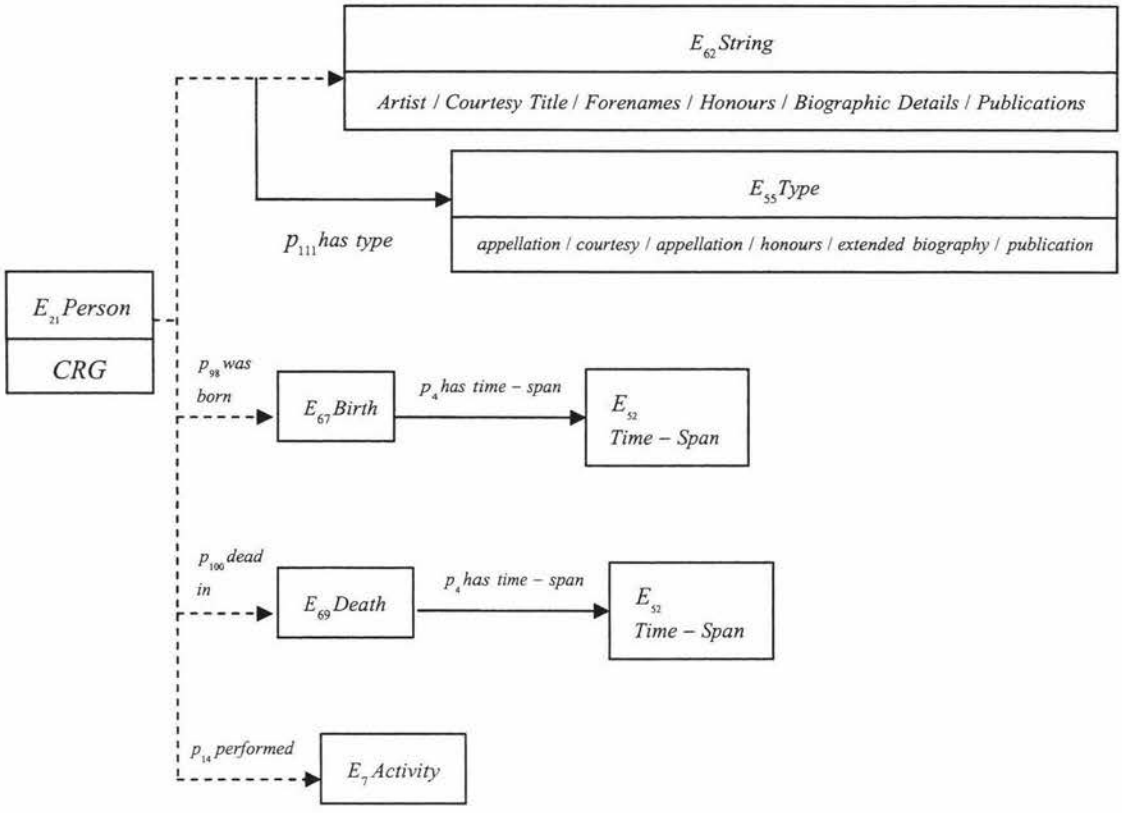
6.2.3.8 Mapping for "Who made it?"

Creator Group is the attribute used to group the fields associated with the creator of the art work. Same mapping as CRG in AMICO, it can be mapped to $E_{21}Person$ or mapped to $E_{39}Actor$ in general. Refer to CRG mapping for details.

The following fields are those subfields covered by Creator Group:

- Artist is the name of the creator. It can be mapped as $E_{62}String$.
- Courtesy Title refers to the courtesy title of the artist. It maps to $E_{62}String$ as well.
- Forenames refer to the surname of the artist. It maps to $E_{62}String$.
- Honours refer to a free text description of the awards have been entitled to the artist. It maps to $E_{62}String$.
- Date of Birth represent the date of the birth of the creator, it is a linked to the $E_{67}Birth$ event. It maps to $E_{52}Time-Span$, mapping of Date of Birth is the same as the mapping of CBD in AMICO.
- Date of Death represent the death date of the creator, similar with Date of Birth, it maps to $E_{52}Time-Span$.
- Nationality is the culture or nationality of the creator of the artwork. Same as with the mapping of CRC in AMICO, the Nationality can be mapped to $E_{55}Type$, $E_4Period$ or $E_{74}Group$. More details can be obtained by refering to the CRC mapping.
- Biographical Details is a text description of the biography for the creator of the art work, it maps to $E_{62}String$ in CRM.
- Chronology is a description of the creator's arranged activity listed in time. It maps to $E_{62}String$.
- Publications refer to a text description of the article published by the creator. It maps to $E_{62}String$.
- Artist Exhibitions refers to the exhibitions of the artist's work. It can be mapped as $E_7Activity$.

Mapping diagram:



Notation:

$$[\![\langle p_3, p_{111}\rangle^6, \langle p_{98}, p_4\rangle, \langle p_{100}, p_4\rangle, p_{14}]\!] : E_{21} \rightarrow \square\left(\left(E_{62} \times E_{55}\right)^5, E_{67} \times E_{52}, E_{69} \times E_{52}, E_7\right)$$

6.2.3.9 Mapping for "What is it related to?"

- Related Document Group is the group contains the fields documenting related work of art, it corresponds to $E_{22} Man-Made\ Object$ in CRM. The fields within the Related Document Group are: Reference and Notes.
  - Reference is a description of relationship between this art work and the others.
  - Notes contains any other information of the artwork which does not readily fit into any other field.

  Both Reference and Notes map to $E_{62} String$.

  Notation of the mapping of Related Document Group:

  $$[\![\langle p_3, p_{3.1}\rangle^2]\!] : E_{21} \rightarrow \square\left(\left(E_{62} \times E_{55}\right)^2\right).$$

- Related Image Group is the group fields documenting related images. Related Multimedia Group maps to $E_{38} Image$.

Notation:

$[\![\langle p_{67}, p_{67}\rangle]\!] : E_{22} \rightarrow \square(E_{28} \times E_{38})$.

Related Image Group contains fields documenting related images, these fields are: Photo Reference and Picture.

Photo Reference refers to the image of the artwork itself, and Picture refers to any known illustration of a work in a publication. Both of them can be mapped to $E_{38}Image$.

- Location refers to the place where the current work is stored. It can be mapped as $E_{53}Place$, the mapping can be:

  $SCR.Location = E_{22}Man-Made\ Object : p_{55}has\ current\ location : E_{53}Place$.

  Notation:

  $[\![p_{55}]\!] : E_{22} \rightarrow \square E_{53}$.

- Price Group is the group contains the fields related with the value of the art work, these fields include: Purchase Price, Valuation for Insurance and Loans. Because all these fields are related with the price, they all map to $E_{62}String$ in CRM.

  Notations for these three fields are the same as:

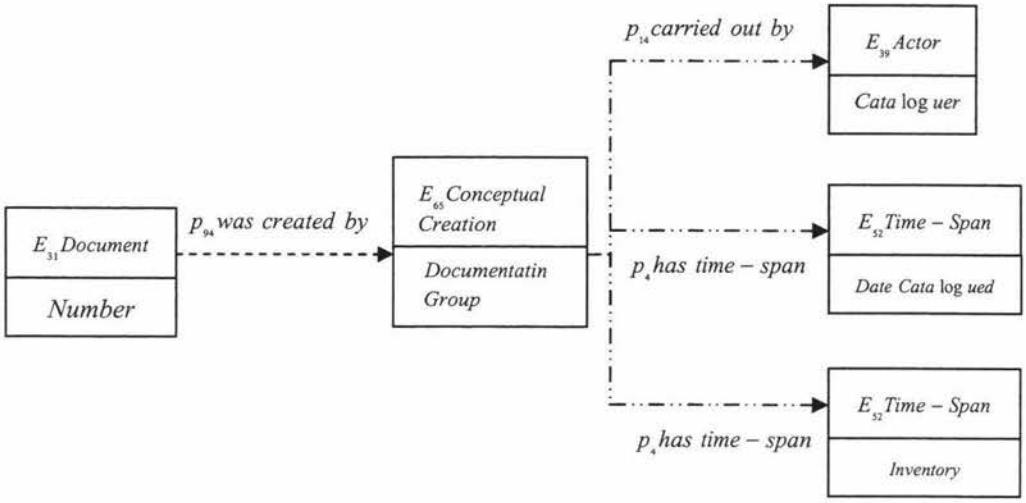
  $[\![\langle p_3, p_{111}\rangle]\!] : E_{22} \rightarrow \square(E_{62} \times E_{55})$.

6.2.3.10 Mapping for "Who documented it?"

- Documentation Group contains the fields recording the documentation history of the art work, and recording the inventory of the collection. It comprises the fields of: Cataloguer/Data Entry, Date Catalogued and Inventory. Similar to AIC in AMICO, Suter dataset object maps to an instance of $E_{31}Document$, so Documentation Group maps to $E_{65}Conceptual\ Creation$ event that associated with the Suter Number related document.
  - Cataloguer/Data Entry refers to the person who did the data entry of the catalogue data, it maps to $E_{39}Actor$.
  - Date Catalogued refers to the data that the catalogue data has been input. It maps to $E_{52}Time-Span$.
  - Inventory refers to the updated by physical sighting of a work and record the date the object was last seen on. It maps to $E_{52}Time-Span$ as well.

Following is the diagram shows the mapping of Documentation Group and its related fields:

Mapping diagram:



Notation:

$$[\![p_{94}[p_{14}, p_4, p_4]]\!] : E_{31} \rightarrow \square\left(E_{65} \times \left(E_{39} \oplus E_{52} \oplus E_{52}\right)\right).$$

6.2.3.11 Mapping for "Multimedia Metadata Fields"

As stated previously that Suter Gallery keeps numbers of multimedia files of the work collection, which includes photographs, negatives and transparencies. Following are the two fields of these multimedia files:

- Photographer: indicates the name of the person who created the multimedia files: *Photographer = DC.Creator.PersonalName*
- Medium/Index.: refers to the numbering system for photographs/negatives/transparencies of collection items:
  Medium / Index = DC.Resource.Identifier

It is the same with the mapping of AMICO multimedia files, they can be mapped to the corresponding fields in Dublin Core, refer to the mapping of *DC.Creator.PersonalName* and *Resource.Identifier* in the corresponding AMICO category for more details.

6.3 Mapping of Te Manawa Data Model to CRM Version 3.0

6.3.1 Introduction of Te Manawa Collection and Data Model

Te Manawa Museum, Galley and Science Centre is based in Palmerston North. It is committed to collecting work by New Zealand artists who have been or who are important to the development of New Zealand art. Te Manawa's collections number around 55,000 items, and include artworks, interactive science exhibits, taonga, heritage objects and natural history specimens (refer to Appendix 5)

The Te Manawa database system was installed by Vernon Systems in 2001. It is a relational database environment. Vernon Systems COLLECTION was designed and developed by Vernon Systems Limited, a New Zealand software company, which has been exclusively dedicated to COLLECTION since 1985. Based on the COLLECTION system, files of Te Manawa's collection have been distributed across several databases, they are: Object, Person, Documentation, Photo/Audio-visual, Event and Other Authority Files. The following description for these file databases was provided by Vernon System Ltd. (Vernon Systems Ltd., 2002).

- *Object is the main file database in Te Manawa's system, it contains all the records belonging to objects in the collection and is comprised of fields such as Accession Number, Name/Title, Acquisition Source, Dimensions, etc; fields that logically describe objects.*
- *Person is a database designed to handle details of both historic and contemporary people and companies and groups of people such as tribes.*
- *Documentation is a database of supportive documentation. It includes fields such as Publisher, Author, Publication Date; it also covers the information related to the updated of documents and the cataloguers.*
- *Photo/Audio-visual is a database of photographs and digital images, motion video and sound.*
- *Event is a database of significant happenings. It is comprised of fields such as the Type of Event, Time/Date and places.*
- *Other Authority Files include files of Place, Classification, Object Status, etc, and its purpose is to provide additional information about Objects, Persons, Documentation and Events.*

A fully described Te Manawa data record field is shown in Table 5 below, the description of each of its data fields is based on Vernon Cataloguing Fields.

Table 5 Te Manawa Museum Data Fields

Te Manawa Fields	Group	Definition
Unique Identification		
TMCR (Te Manawa Category Record)		Refers to the instance of the Suter object.
System ID	N	A unique identifier, assigned to a work in the Te Manawa record.
Object		
Accession No.	N	Used to record an identifier by which an object is usually known and registered.
Other ID	Group	Used to group the associated fields documenting object numbers or identifying data other than those identifiers already catered for by specific fields.
ID	Other ID	Used to record object numbers, such as former accession numbers, catalogue numbers,etc...
Type	Other ID	Used to specify the type of ID entered in the associated Other ID field.
Name/Title	N	Used to record the name or title of an object or work of art.
Current Owner	N	Used to record the name of the current legal or formal owner of an object.
Item Count (default to 1)	N	Used to record the number of discrete and separable pieces which make up one record.
Parts	Group	Used to group the associated fields documenting the parts of an object based on their potential for separate movement.
Part ID	Parts	Used to identify the parts of an object.
Part Name	Parts	Used to describe the part name of an object.
Acquisition	Group	Used to group the associated fields documenting acquisition of the object.
Acquisition Date	Acquisition	Used to record the date on which an object was legally or officially acquired.
Acquistion Method	Acquisition	Used to identify the method with which an object was acquired.
Acquisition Source	Acquisition	Used to record the name of the person or company that transferred custody of an object to your institution.
Acquisition Source Role	Acquisition	Allows people to record the sub-role of the person or company from whom an object was required.
Acquisition Notes	Acquisition	Used to record any comments about acquisition of an object that are not specifically catered for in other fields.
Acquistion Price Local	Acquisition	Used to record the purchase price of an object in the local currency.
Brief Description	N	Used to record a general description of an object.
Media/Materials	Group	Used to record associated fields recording precise and individual details about the materials or media of which an object is made.
Media/Materials Desciption	Media/Materials	Used to record a text description of the materials or media of which an object is made.
Measurement	Group	Used to group fields recording measurements.
Measurement Desc	Measurement	Used to record a text description of the measurements of an object.
Measurement Type	Measurement	Used to indicate the type of measurement being recorded for an object.
Measurement Reading	Measurement	Used to record the measurements of an object.
Measurement Notes	Measurement	Used to record any comments about measurement are not specifically catered for elaswhere.
Condition	Group	Used to group the associated fields documenting the condition of the object.
Condition Person	Condition	Used to record the name of the person responsible for condition details.
Condition Date		Used to record the date of the condition recorded.
Condition Keywords	Condition	Used to record a word or brief phrase that summarizes the condition of an object.
Condition Notes	Condition	To record any comments about the condition of an object are not specifically catered elsewhere.
Treatment	Group	Used to group the associated fields documenting the treatment record.
Treatment Description	Treatment	Used to record a description of the treatment carried out on an object.
Treatment Person	Treatment	Used to record the name of the person responsible for Treatment details.
Treatment Date	Treatment	Used to record the date of the treatment.
Treatment Notes	Treatment	Records any comments about the treatment of an object are not specifically catered elsewhere.
Signature/Marks	Group	Group the associated fields documenting the signature and date or the marks on the object.
Signature/Marks	Signature/Marks	Used to record the exact text of any signature of inscription.
Signature/Marks Type	Signature/Marks	Used to specify the type of mark applied to an object.
Signature/Inscription Method	Signature/Marks	Used to specify the method used to apply a mark or inscription to an object.
Signature/Marks Notes	Signature/Marks	Used to record comments about the signature or marks are not specifically catered elsewhere.

Te Manawa Fields	Group	Definition
Object		
Provenance	Group	Used to group the associated fields documenting the provenance details.
Provenance Date	Provenance	Used to record the principal date associated with each aspect an object's provenance.
Provenance Details	Provenance	Used to record information about the history of an object.
Provenace Person	Provenance	Used to record the name of the primary person or company associated with each aspect an object's provenance.
Provenance Place	Provenance	To record the name of the pirmary place assoicated with each aspect an object's provenance.
Primary Maker	Group	Used to group the associated fields documenting the person or company primarily responsible for the creation of an object.
Primary Maker	Primary Maker	Used to describe the name of the primary maker.
Primary Maker Role	Primary Maker	Allows people to record the sub-role of the person or company primarily responsible for the creation of an object.
Primary Production Date	Primary Maker	Used to record the date on which an object was made.
Primary Production Place	Primary Maker	Used to record the name of the place an object was made.
Person		
Person	Group	Used to group the fields associated with the person exsiting in the file system.
Biographical Details	Person	Records all the detailed information about the person, it includes person type, corporate type, name, gender, ethnicity, nationality, etc...
Documentation		
Documentation	Group	Used to group the fields recording the documentation history of the work.
Cataloguer	Documentation	Used to record the names of people who subsequently modify the record for the object.
Cataloguer Date	Documentation	Used to record the date an object was intially catalogued and/or subsequent dates on which the record was modified.
Photo/Audio-Visual		
Photo/Audio-Visual Reference	N	Used to record an identifying reference for an image documenting an object in the collection. It is usually a number, but could be any data that will assist people identiy the image.
Photo/AV Notes	N	It is a narrative description of the contents of the media file.
Event		
Related Activities	Group	Used to group the fields recording the related acitivites of the object.
Related Exhibition Venue	Related Activities	Used to record the venues of any exhibitions an object has been included in.
Exhibition Venue Details	Related Activities	Used to record the details of any exhibitions an object has been included in.
Related Outward Loan	Related Activities	Used to record any exhibitions or outward loans an object has been included in.
Outward Loan Details	Related Activities	Used to record details about any exhibitions or outward loans an object has been included in.
Other Authority Files		
Record Status	N	Used to indicate the completeness or correctness of a record.
Record Status History	N	Used to indicate the history of the record status for the current record.
Department	N	Used to idicate which curatorial department within the insitution are responsible for an object.
Collection	N	Used to identify a named collection of which an object is part.
Classification	N	Used to categorize an object by assigning it to a group or set of like items.
Credit Line	N	Used to record the approved text that should be associated with an object whenever it is displayed, published or otherwise presented to the public.
Fund	Group	Used to group the fields related with funding.
Funder	Fundiing	It is a free text description of the person or company who funded the object.
Funding Type	Fundiing	It indicates the type of the funding.
Amount Funded	Fundiing	Refers to the amount has been funded for the object.
Disallow Movement?	N	Indicate whether the object is moveable,
Current Location	Group	Used to group the fields associated with the current location of the object.
Location Reason	Current Location	Used to defines why an object is in that location.
Location: Current	Current Location	Shows the place the object is currently located.
Location:Date	Current Location	Records the date on which a Movement or Inventory Transaction took place.
Associated Person	N	Used to record the name of a person or company associated with an object, such as a person who used an object.

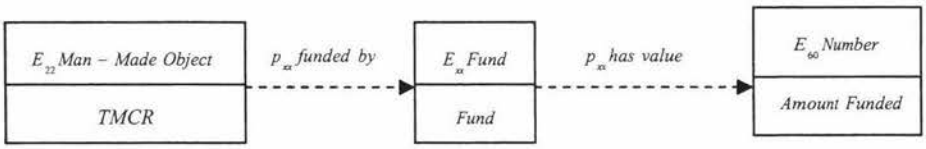
6.3.2 Mapping Formalism

As with the mapping of AMICO data and the Suter data, the mapping for Te Manawa is also based on CRM Version 3.0. The mapping formalism follows that adopted in the few process validation exercises in this thesis. Refer to section 5.1.2.2.

The Te Manawa fields can be mapped to the corresponding entity in CRM, such as: $Parts = E_{18}Physical\ Staff$, and its correspondence of the relation expressed by the field to a CRM link can be written as:

$TMCR.Parts = E_{22}Man-Made\ Object : p_{46}is\ composed\ of : E_{18}Physical\ Stuff$ (TMCR represent the Te Manawa catalogue as a whole, this will be explained in the following sections)

If the relation expressed by a Te Manawa field relates to a path with intermediate entities in the CRM, it is a "join" relationship, a diagram will be used to illustrate the relationship, see the example below:



6.3.3 Mapping Te Manawa Data to CRM

The mapping of Te Manawa is not unlike the mapping of AMICO data and Suter data. The mapping of Te Manawa data can be regarded as two semantic levels: the first level can be the completed record for an object; it can be treated as an instance of $E_{31}Document$, which can be identified by 'System ID' and its 'Accession Number' as the data identifier. For example, documents about the related object have been sorted in different file folders in a filing cabinet system, and there is a tag number attaching to each of the file folders that indicating the location of those documents. So the tag number is similar to the 'System ID' and 'Accession Number', which represents the whole set of documents. The second level is the actual content of the Te Manawa dataset about an object in terms of its data fields to which it refers and relates.

6.3.3.1 Mapping for Unique Identification

- Each of the Te Manawa catalogue records is in one-to-one correspondence with the described object, Te Manawa Catalogue Record (TMCR), corresponding to that in the CRM can be expressed as:

  $TMCR = E_{22} Man-Made\ Object$

- As mentioned above, the filed System ID of the Te Mamawa dataset object can be mapped as an instance of $E_{31} Document$ that used to identify the related object content, so there is:

  $SystemID.TMCR = E_{31} Document : p_{70} documents : E_{22} Man-Made\ Object .$

  Notation:

  $\langle p_{70} \rangle : E_{31} \rightarrow E_{22} .$

6.3.3.2 Mapping for "Object"

- Accession Number is the number that is designated to the work itself. The mapping is the same with the 'Accession Number' in Te Manawa, it maps to $E_{42} Object\ Identifier$. Refer to mapping details in the 'Accession Number' of Suter for more information

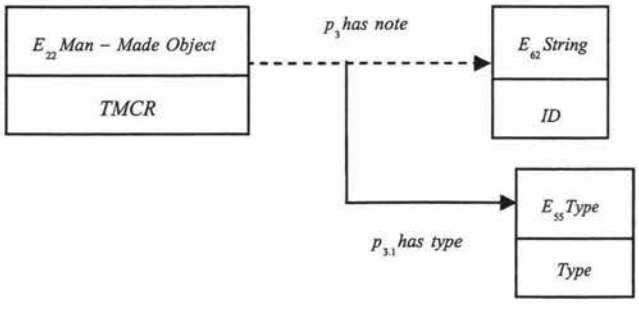
  Notation:

  $\langle p_{47} \rangle : E_{22} \rightarrow E_{42} .$

- Other ID is a group with all the other format of IDs defined by Te Manawa for certain identification purpose. 'Other ID Group' contains the fields: ID and Type.
  - ID is the identification number, it can be mapped as $E_{62} String$.
  - Type indicates the specific type of the ID, which can be mapped as $E_{55} Type$ attaches to $E_{62} String$.

  Refer to the following mapping diagram for Other ID Group:

  Mapping diagram:



Notation:

$[\![\langle p_3, p_{3.1}\rangle]\!]: E_{22} \rightarrow \square(E_{62} \times E_{55})$.

- Name/Title refers to the name or the title of the object, same with the mapping of 'Work Title' in Suter field, Name/Title maps to $E_{35}Title$, refer to 'Work Title' mapping for more details.
- Current Owner refers to the current owner of the object, it maps to $E_{39}Actor$, and the mapping is:

  $TMCR.Current\ Owner = E_{22}Man-Made\ Object: p_{52}has\ current\ owner: E_{39}Actor$.

  Notation:

  $[\![p_{52}]\!]: E_{22} \rightarrow \square E_{39}$.
- Item Count refers to the number of pieces of the object. The mapping is similar to the mapping of 'No. of pieces' in Suter fields. Refer to 'No. of pieces' mapping in Suter for more details.

  Notation:

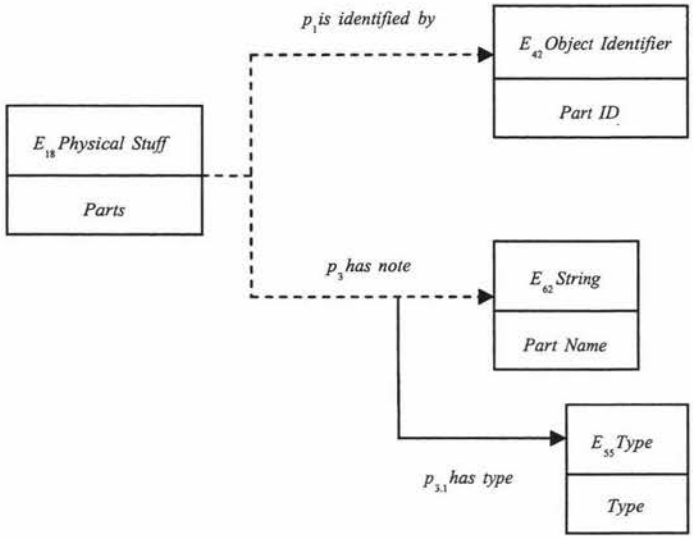
  $[\![p_{46}\langle p_{57}\rangle]\!]: E_{22} \rightarrow \square(E_{22} \times E_{60})$.
- Parts used to group the fields documenting all the relevant parts attaches to the object. Parts can be mapped as $E_{18}Physical\ Staff$, and its relationship with Te Manawa Catalogue Record (TMCR) is:

  $TMCR.Parts = E_{22}Man-Made\ Object: p_{46}is\ composed\ of: E_{18}Physical\ Stuff$.

  The fields covered by Parts are: Part ID and Part Name. Part ID represents the identify number of the relevant part and Part Name refers to the name of the relevant part.

  The mapping of Parts and its fields is showed in the following diagram:

Mapping diagram:



Notation:

$[\![p_1, \langle p_3, p_{3.1} \rangle]\!] : E_{18} \rightarrow \square(E_{42}, E_{62} \times E_{55})$.

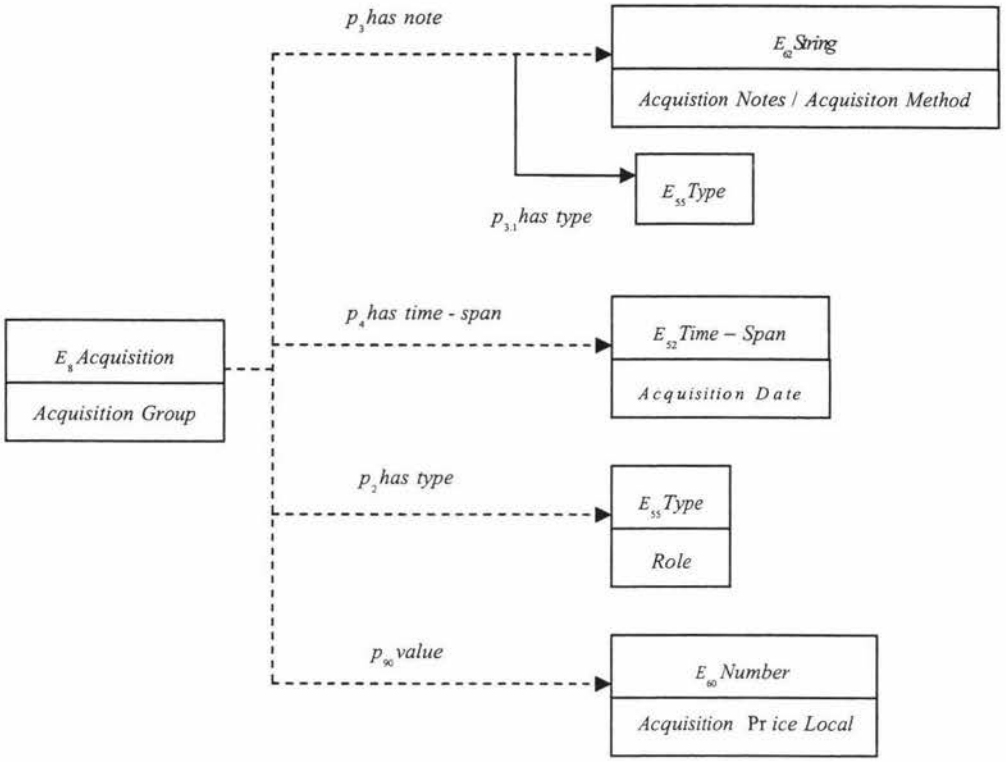
- Acquisition Group is used to group the fields associated with the acquisition of the object. These fields are: Acquisition Date, Acquisition Method, Acquisition Source, Acquisition Source Role, Acquisition Notes and Acquisition Price Local.

  Acquisition Group can be mapped as $E_8 Acquisition$, the mapping is the same as the mapping of 'Acquisition Group' in Suter. Refer to Suter for more mapping details.

  - Acquisition Date maps to $E_{52} Time-Span$.
  - Acquisition Method identifies the method with which the object was acquired. It maps to $E_{62} String$.
  - Acquisition Source indicates the source of the object. It can be mapped as $E_{39} Actor$.
  - Acquisition Source Role maps to $E_{55} Type$.
  - Acquisition Notes is a free text description of the acquisition. It can be mapped as $E_{62} String$.
  - Acquisition Price Local refers to the purchase price of the object, it maps to $E_{60} Number$.

  Refer to the following mapping for Acquisition Group:

Mapping diagram:



Notation:

$[\![\langle p_3, p_{3.1}\rangle^2, p_4, p_2, p_{90}]\!] : E_{22} \rightarrow \square\left((E_{62} \times E_{55})^2, E_{52}, E_{55}, E_{60}\right).$

- Brief Description is used to record a general description of an object. Same with the mapping of 'Physical Description' in AMICO, Brief Description can be mapped as $E_{62}String$, refer to 'Physical Description' in AMICO for more mapping details.
- Media/Materials Group is used to group fields documenting materials or media of which an object is made. The subfields of Media/Materials Group includes: Media/Materials Description.

  Media/Materials Group can be mapped as $E_{12}Production$:

  $Media\ /\ Materials = E_{12}Production$.

  When Media/Materials Gorup links with Te Manawa Catalogue Record (TMCR), its relationship can be described as:

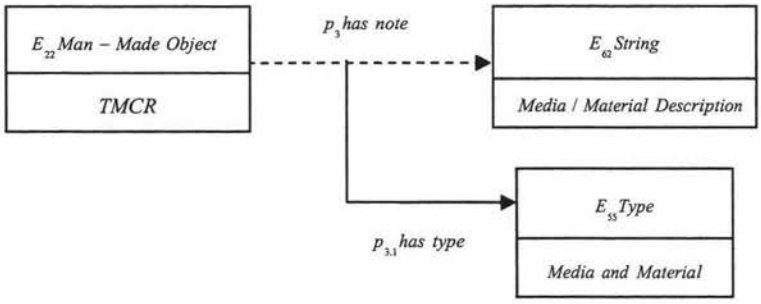
  $TMCR.Media\ /\ Material = E_{22}Man-Made\ Object : p_{31}was\ produced\ by : E_{12}Production$

  Notation:

  $[\![p_{31}]\!] : E_{22} \rightarrow E_{12}$.

- Media/Material Description is a free text description of the materials and techniques used to create the work of art. It maps to $E_{62}String$, and the mapping diagram is:

  Mapping diagram:

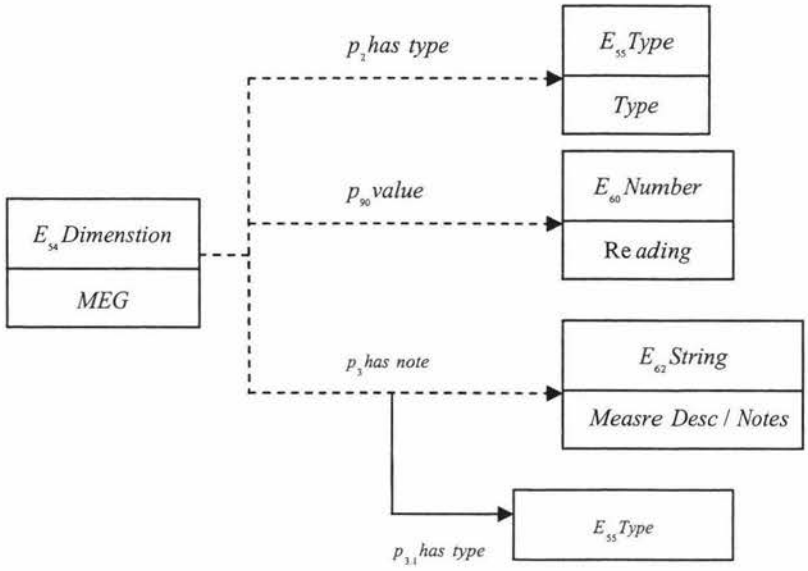


  Notation: $[\![\langle p_3, p_{3.1} \rangle]\!] : E_{22} \rightarrow \square(E_{62} \times E_{55})$

- Measurement Group used to group fields recording measurements. These fields are: Measurement Desc, Measurement Type, Measurement Reading and Measurement Notes. Same as the mapping of Measurement Group in Suter, it can be mapped to $E_{54}Dimension$. Refer to the mapping of Measurement Group in Suter for more details.
  - Measurement Desc is a brief description of the measurement of an object. It can be mapped to $E_{62}String$.
  - Measurement Type is a term indicating the type of measurement being taken. It can be mapped to $E_{55}Type$.
  - Measurement Reading used to record the measurement of an object, it can be mapped to $E_{60}Number$.
  - Measurement Notes is a free text description. It maps to $E_{62}String$.

  Mapping of Measurement Group and its related fields is shown in the following diagram:

Mapping diagram:



Notation:

$[\![p_2, p_{90}, \langle p_3, p_{3.1} \rangle]\!] : E_{54} \rightarrow \square(E_{55}, E_{60}, E_{62} \times E_{55})$.

- Condition Group used to group the fields recording the condition of the object. These fields include: Condition Person, Condition Date, Condition Keywords and Condition Notes.

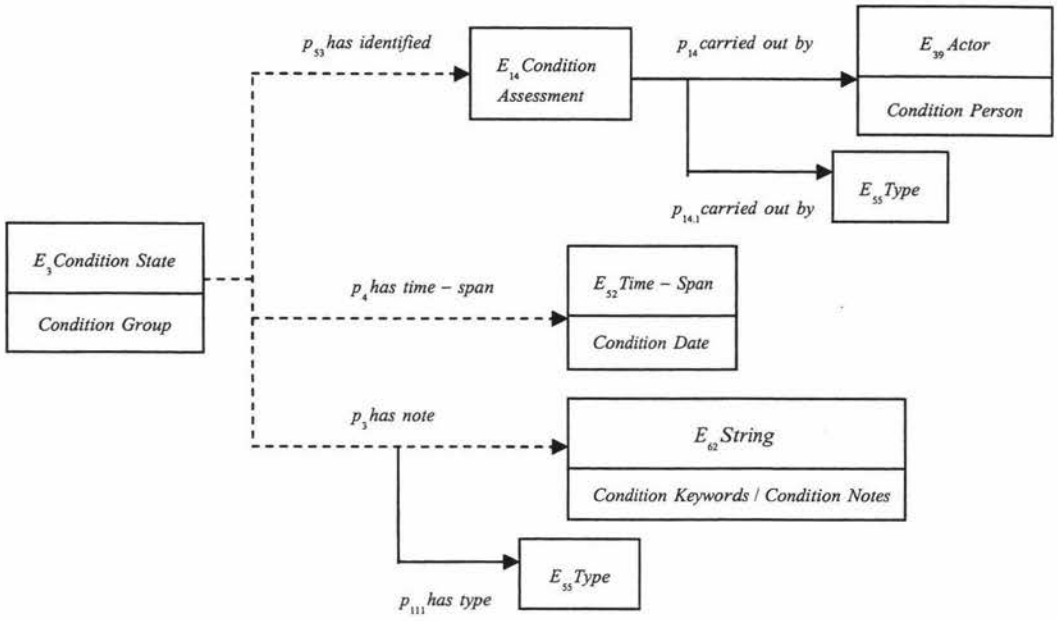
  Condition Group can be mapped as $E_3 Condition\ State$, its relationship with Te Manawa Catalogue Record (TMCR) is:

  $TMCR.Condition\ Group = E_{22} Man-Made\ Object : p_{44} has\ condition : E_3 Condition\ State$

  - Condition Person refers to the name of the person responsible for condition details. It maps to $E_{21} Person$.
  - Condition Date used to record the date of the condition recorded, it maps to $E_{52} Time-Span$.
  - Condition Keywords is a free text description of the overall condition of the object. It maps to $E_{62} String$.
  - Condition Notes used to record any comments about the condition of an object are not specifically catered elsewhere. It maps to $E_{62} String$ as well.

  See the diagram below for the mapping of Condition Group and its fields:

Mapping diagram:



Notation:

$$[\![p_{53}\langle p_{14}, p_{14.1}\rangle, p_4, \langle p_3, p_{3.1}\rangle^2]\!] : E_3 \rightarrow \square(E_{14} \times (E_{39} \times E_{55}), E_{52}, E_{62} \times E_{55}).$$

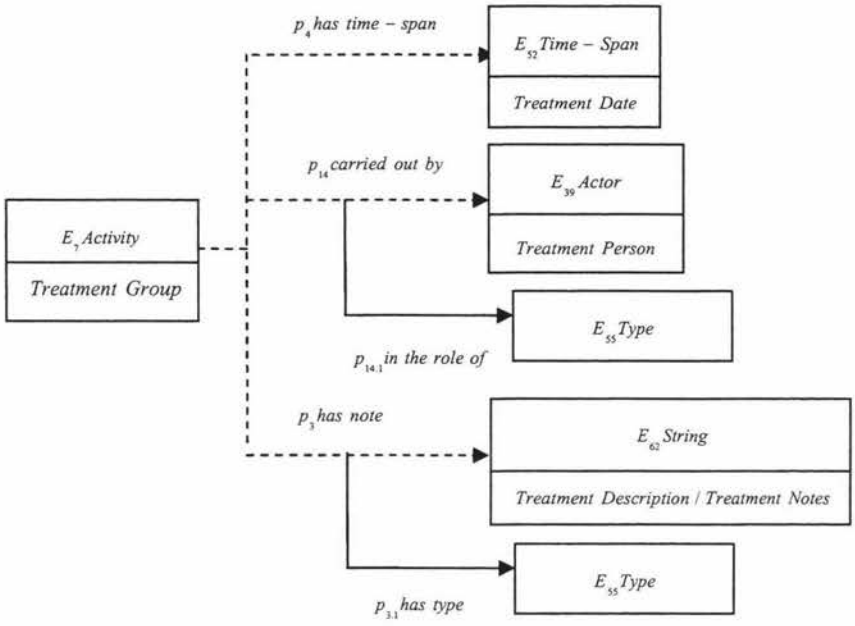
- Treatment Group is used to group the fields recording the treatment for the object. These fields include Treatment Description, Treatment Person, Treatment Date and Treatment Notes. Treatment Group can be regarded as an activity and so it can be mapped as $E_7 Activity$. Its associated relation with Te Manawa Catalogue Record (TMCR) is:

  $E_{22} Man-Made\ Object : p_{16} used\ object : E_7 Activity$.

  - Treatment Description is the description of the treatment carried out on an object. It maps to $E_{62} String$.
  - Treatment Person is the name of the person responsible for treatment details. It maps to $E_{39} Actor$.
  - Treatment Date is used to record the date of the treatment. It maps to $E_{52} Time-Span$.
  - Treatment Notes is the free text descriptions of the treatment of an object are not specifically catered elsewhere. It maps to $E_{62} String$.

  Following is the mapping diagram for Treatment Group and its related fields:

Mapping diagram:



Notation:

$[\![p_4, \langle p_{14}, p_{14.1} \rangle, \langle p_3, p_{3.1} \rangle]\!] : E_7 \rightarrow \square(E_{52}, E_{39} \times E_{55}, E_{62} \times E_{55})$.

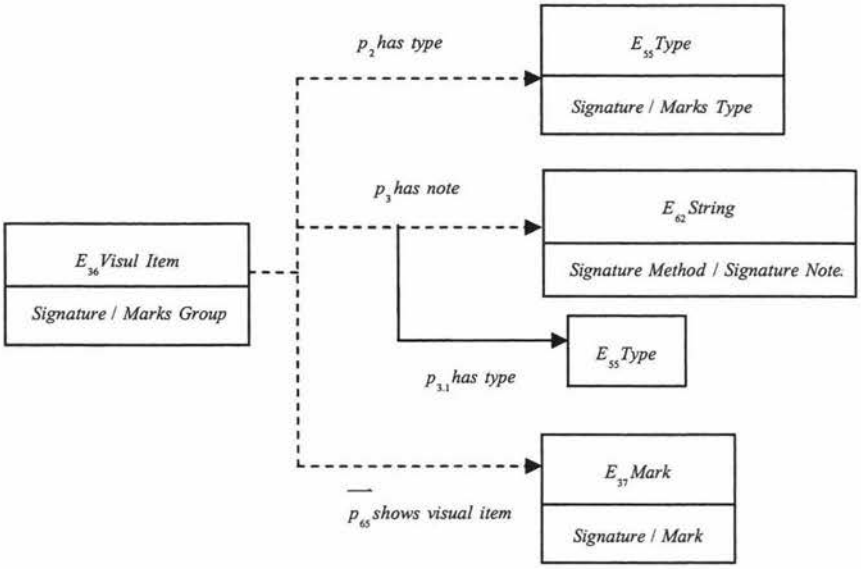
- Signature/Marks Group used to group the fields associated with the signature and inscription of the object. These fields are: Signature/Marks, Signature/Marks Type, Signature/Inscription Method, and Signature/Marks Notes.

  The Signature/Marks Group can be mapped to E_{36}*Visual Item* , and its associated relationship with Te Manawa Catalogue Record (TMCR) is:
  TMCR.Signature / Mark Group =
  E_{22}*Man – Made Object* : p_{65}*shows visual item* : E_{36}*Visual Item*

  - Signature/Marks used to record the exact text of any signature or inscription. It maps to E_{37}*Mark* .
  - Signature/Marks Type refers to the type of mark applied to an object, it maps to E_{55}*Type* .
  - Signature/Inscription Method refers to the method used to apply a mark or inscription to an object. It maps to E_{62}*String* .
  - Signature/Marks Notes refer to the text description of the signature. It maps to E_{62}*String* .

  Diagram below shows the mapping of Signature/Mark Group and its related fields:

Mapping diagram:



Notation:

$[\![p_2, \langle p_3, p_{3.1} \rangle, \overline{p_{65}}]\!] : E_{36} \rightarrow \square(E_{55}, E_{62} \times E_{55}, E_{37})$.

- Provenance Group used to group the fields associated with the fields describe history of past owners of the object. These fields are: Provenance Date, Provenance Details, Provenance Person and Provenance Place.

  Provenance Group itself can be treated as activity, so it can be mapped as $E_7 Activty$, its relationship with Te Manawa Catalogue Record (TMCR):

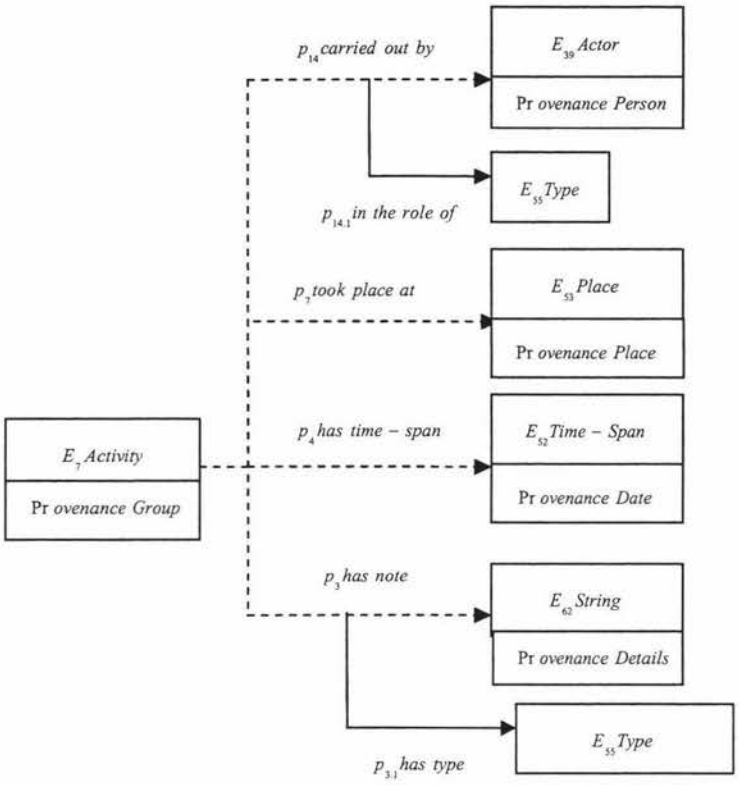
  $TMCR.Provenance\ Group =$

  $E_{22} Man-Made\ Object : p_{20} has\ specific\ purpose : E_7 Activity$

  - Provenance Date is used to record the date associated with an object's provenance. It maps to $E_{52} Time-Span$.
  - Provenance Details used to record information about the history of an object, it maps to $E_{62} String$.
  - Provenance Person used to record the name of the person or company associated with an object's provenance, it maps to $E_{39} Actor$.
  - Provenance Place refers to the name of the place associated with an object's provenance. It can be mapped to $E_{53} Place$.

  Following is the diagram shows the mapping for Provenance Group and its fields:

Mapping diagram:



Notation:

$[\![\langle p_{14}, p_{14.1}\rangle, p_7, p_4, \langle p_3, p_{3.1}\rangle]\!] : E_7 \rightarrow \square(E_{39}, E_{55}, E_{52}, E_{62} \times E_{55})$.

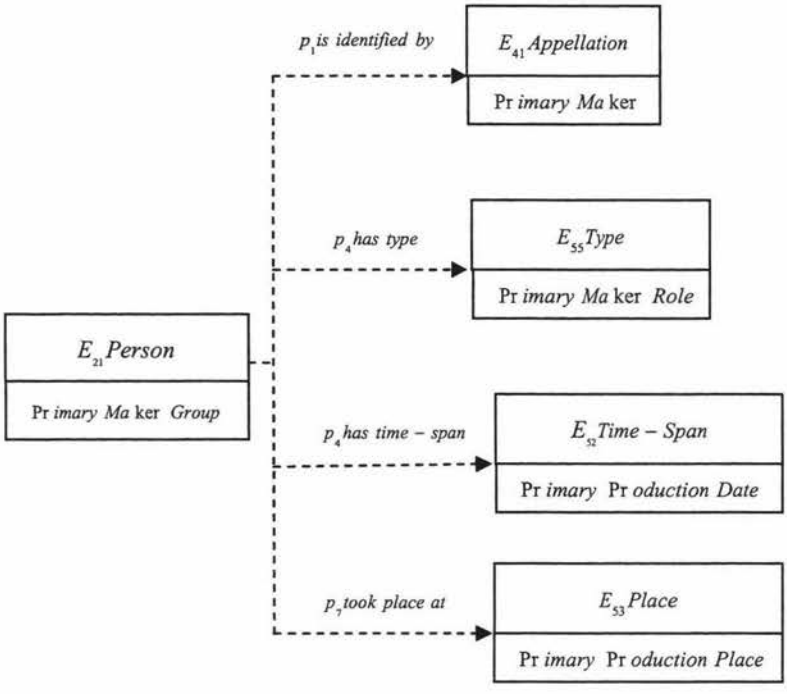
- Primary Maker Group used to group the fields associated with the person or company primarily responsible for the creation of the object. These fields are: Primary Maker, Primary Maker Role, Primary Production Date and Primary Production Place.

  Same with the 'Creator' in Suter, Primary Make Group maps to $E_{21}Person$. Refer to the mapping of 'Creator' in Suter for more details.

  - Primary Maker refers to the name of the primary marker, it maps to $E_{41}Appellation$.
  - Primary Maker Role is the person responsible for the creation of the object. It maps to $E_{55}Type$.
  - Primary Production Date refers to the date on which an object was made, it maps to $E_{52}Time-Span$.
  - Primary Production Place records the name of the place an object was made, it maps to $E_{53}Place$.

  The diagram below shows the mapping of Creator Group and its related fields:

Mapping diagram:



Notation:

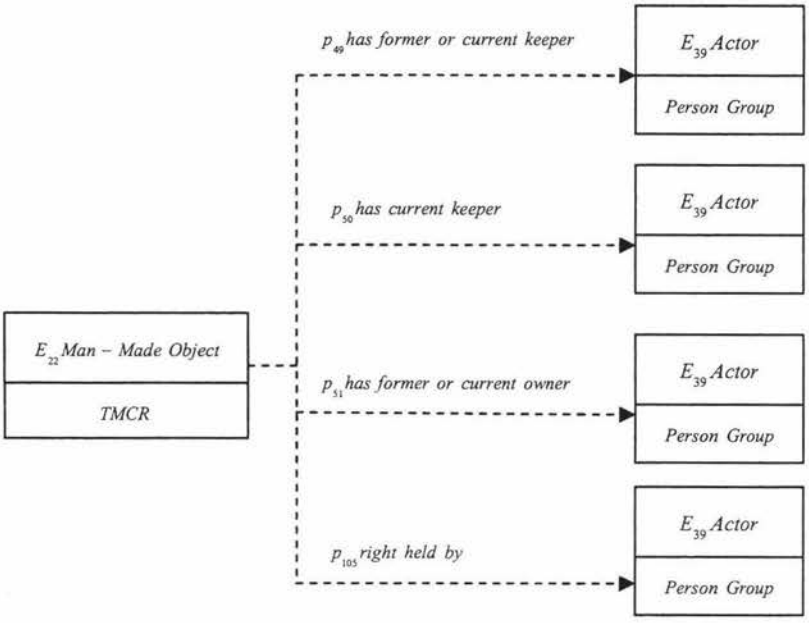
$[\![p_1, p_2, p_4, p_7]\!] : E_{21} \rightarrow \square(E_{41}, E_{55}, E_{52}, E_{53})$.

6.3.3.3 Person

- Person Group used to group the person details for both historic and contemporary people and companies existing in Te Manawa data files. Person Group contains field Biographical Details. Person Group maps to E_{39} *Actor*.

  Person Group's associated relationship with Te Manawa Catalogue Record (TMCR) can be treated as:

Mapping diagram:



Notation:

$[\![p_{49}, p_{50}, p_{51}, p_{105}]\!] : E_{22} \rightarrow \square(E_{39}, E_{39}, E_{39}, E_{39})$.

- Biographical Details also covers the details of role, gender, name, first names, last name, initials, life years, year born, year died, age, life date notes, place of birth, these are all text description, so Biographical Details map to $E_{62} String$:

$$Person\ Group = E_{22} Man - Made\ Object : p_3 has\ \text{note}^{[p_{3.1} has\ type]} : E_{62} String$$

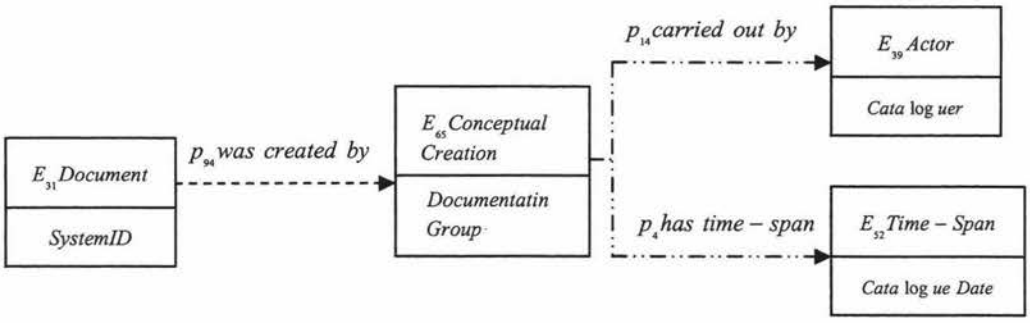
Notation:

$[\![\langle p_3, p_{3.1} \rangle]\!] : E_{22} \rightarrow \square(E_{55} \times E_{62})$.

6.3.3.4 Documentation

- Documentation Group is used to group the fields recording the document history of the work. These fields are: Cataloguer, Catalogue Date. Same as the 'Documentation Group' in Suter, it can be mapped to $E_{65} Conceptual\ Creation$ event that associated with the SystemID related document.
  - Cataloguer refers to the names of people who subsequently modify the record for the object, it maps to $E_{39} Actor$.
  - Catalogue Date refers to the data that the catalogue data has been input. It maps to $E_{52} Time - Span$.

Following is the diagram shows the mapping of Documentation Group and its related fields:



Notation:

$$[\![p_{94}[p_{14},(p_4)^2]]\!]: E_{31} \rightarrow \square\left(E_{65} \times \left(E_{39} \oplus E_{52}\right)\right)$$

6.3.3.5 Photo/Audio-Visual

- Photo/Audio-Visual Reference is used to record an identifying reference for an image documenting an object in the collection. Same as the mapping of Photo: Medium/Index in Suter, the Photo/Audio-Visual Reference maps to DC.Resource.Identifier in Dublin Core. Refer to Suter for more mapping details.
- Photo/AV Notes is a narrative description of the contents of the media file, it maps to DC.Description in Dublin Core. Refer to Dublin Core for more details.

6.3.3.6 Event

- Related Activities Group used to group the fields related to the exhibition related activities of the object. These fields are: Related Exhibition Venue, Exhibition Venue Details, Related Outward Loan and Outward Loan Details. Related Activities Group maps to $E_7 Activity$. The mapping details are same with the mapping of 'Artwork Exhibition' in Suter; refer to Suter for more details.
  - Related Exhibition Venue describes the venues of related exhibitions for the object.
  - Exhibition Venue Details is the description of the details of the exhibition, includes venue place, venue opening and closing time.
  - Related Outward Loan refers to the borrower's information.

- Outward Loan Details describe the details of the outward loan, which includes the details of the borrowed object.

All the above fields map to $E_{62}String$.

Notation for Related Activities Group is:

$[\![\langle p_3, p_{3.1}\rangle^4]\!]: E_7 \rightarrow \square(E_{62} \times E_{55})^4$.

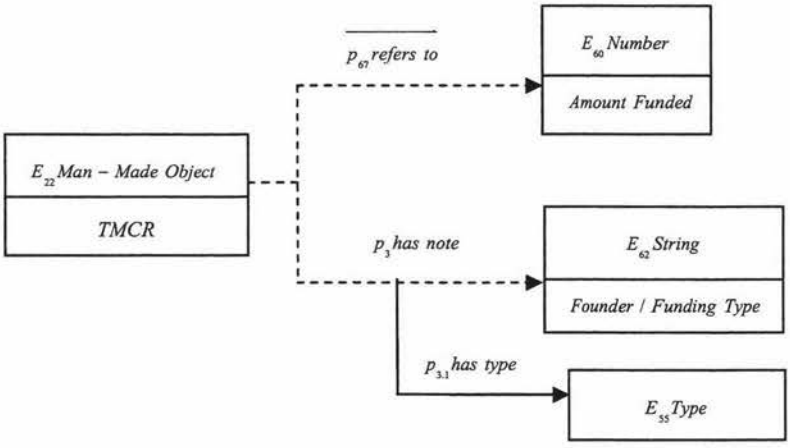
6.3.3.7 Other Authority Files

The 'Other Authority Files' contains an array of files and group.

- Record Status indicates the completeness or correctness of a record.
- Record Status History is a text description used to indicate the history of the record status for the current record, it maps to $E_{62}String$.
- Department used to indicate which curatorial department within the institution are responsible for an object.
- Collection used to identify a named collection of which an object is part.
- Classification used to categorize an object by assigning it to a group or set of like items. Classification can be mapped to $E_{62}String$.
- Credit Line is the approved text that should be associated with an object whenever it is displayed, published or presented to the public, it can be mapped to $E_{62}String$ as well.

  The above six fields all map to $E_{62}String$, which have the same mapping with Biographical Details. Refer to Biographical Details for more mapping details.
- Fund Group used to group the fields related with funding. These fields are: Funder, Funding Type, and Amount Funded. These fields all associated with Te Manawa Catalogue Record (TMCR).
  - Funder is a free text description of the person or company who funded the object. It maps to $E_{62}String$.
  - Funding Type indicates the type of the funding. It maps to $E_{62}String$.
  - Amount Funded refers to the amount has been funded. It can be mapped as $E_{60}Number$.

  The following diagram shows the mapping of Fund Group:

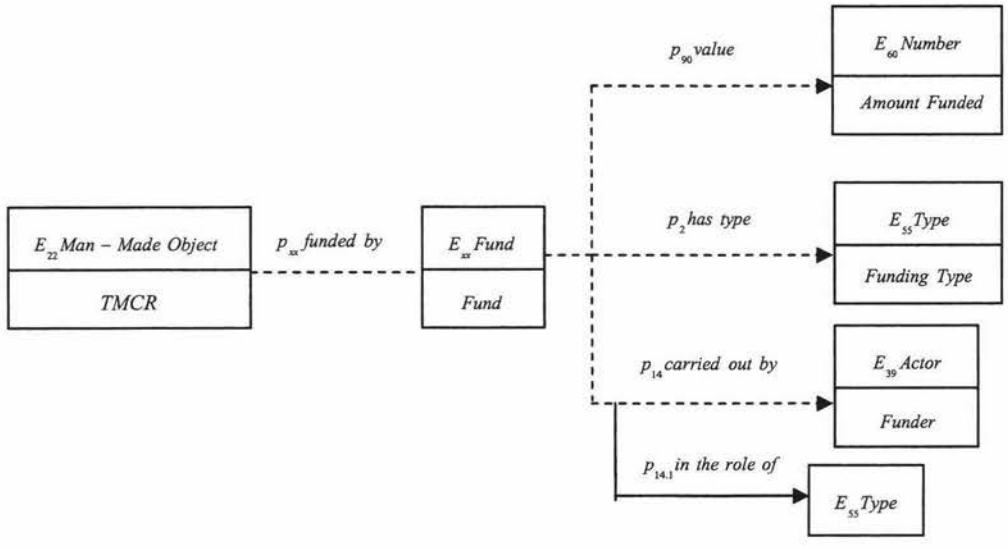


Notation:

$$[\![p_{67}, \langle p_3, p_{3.1} \rangle^2]\!] : E_{22} \rightarrow \square\left(E_{60}, (E_{62} \times E_{55})^2\right)$$

From the mapping of Funding Group, the researcher found that there is an apparent shortcoming in the CRM Version 3.0. The mapping of Funding Group would be more sensible if the CRM Version covered the new entity of Sponsor/Fund with the new property 'sponsor/fund (is sponsored/funded by)' and 'has value'. Because the new entity and the property number are unknown, they can be regarded as $E_{xx}Fund$, $p_{xx}fund(is\ funded\ by)$ and the mapping diagram would become:

Mapping diagram:

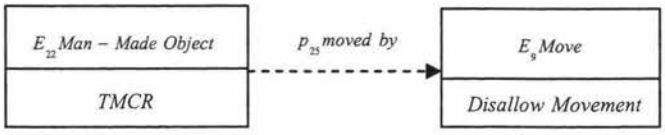


Notation:

$$[\![p_{xx}, [\![p_{90}, p_2, \langle p_{14}, p_{14.1} \rangle]\!]]\!] : E_{22} \rightarrow \square\left(E_{xx} \times \square\left(E_{60}, E_{55}, (E_{39} \times E_{55})\right)\right)$$

- Disallow Movement indicate whether the object is moveable, it can be mapped as $E_9 Move$ in CRM, refer to the following mapping:

Mapping diagram:

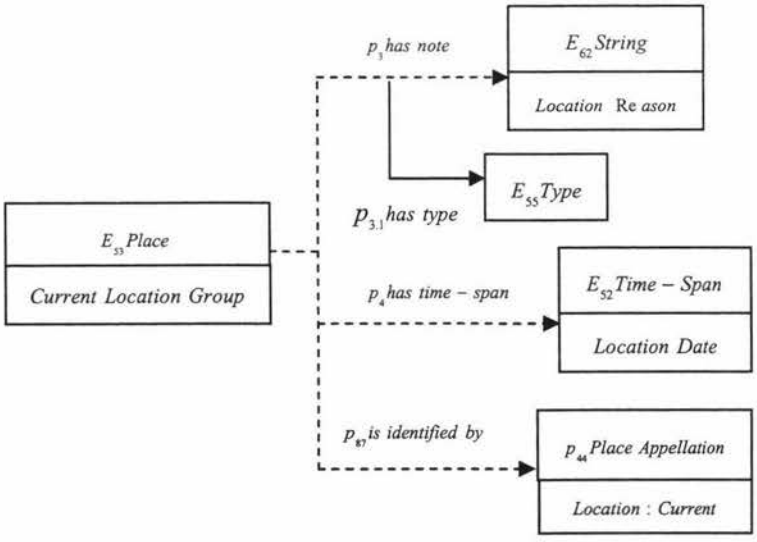


Notation:

$[\![p_{25}]\!] : E_{22} \rightarrow \square E_9$.

- Current Location Group is used to group the fields related with the current location of the object. These fields are: Location Reason, Location: Current and Location Date.
  - Current Location Group can be mapped to $E_{53} Place$, refer to the mapping of 'Location' in Suter for more details.
  - Location Reason indicates the reason for putting the object in the current location, it maps to $E_{62} String$.
  - Location: Current indicates the current location of the object, it maps to $E_{44} Place\ Appellation$.
  - Location Date records the date of the object moved to the current location, it maps to $E_{52} Time-Span$.

  Refer to following mapping diagram for the mapping of Current Location Group and its fields:



Notation:

$[\![\langle p_3, p_{3.1} \rangle, p_4, p_1]\!] : E_{53} \rightarrow \square (E_{62} \times E_{55}, E_{52}, E_{44})$.

- Associated Person refers to the name of a person or company associated with an object, it maps to $E_{39}Actor$, the mapping is similar with the mapping of Person Group in 6.3.3.3, refer to 6.3.3.3 for more mapping details.

6.4 Summary

The validation of the archival information based on the records provided by two New Zealand museums and art galleries (The Suter Gallery in Nelson and Te Manawa in Palmerston North) were successful. Based on these successes the researcher believes the fourth aim of the thesis, "validate the CRM within the New Zealand context", has been achieved.

The two mapping tools, category theory notation and the diagram mapping tool, were able to transfer the content of the archival records to the CRM, while maintaining the richness and quality of the original information. This is further positive confirmation of the second aim of this thesis – "develop tools to assist in the validation of the CRM against real-world heritage collection systems."

The new validation processes once again demonstrates the wide interoperability of the CRM and the compatibility of the CRM with other types of data resource. However, some apparent shortcomings were found during the mapping exercises – these relate to the need to add the extra entity of $E_{xx}Fund$ and a new property $p_{xx}fund(is\ funded\ by)$, these two items are associated with the related 'Loan' information which regarded as an important information to the museum record.

7 Conclusions

The investigation provided a unique opportunity to examine the CIDOC CRM in detail. The adoption of the graphical tool and the application of category theory notation provided further insights in what is a complex and innovative environment. In addition, the investigation highlighted a number of issues relating to the use of the CRM. The category theory notation suggested the possibility of some interesting research as a tool for parsing data from a database to the CRM and subsequently to XML or one of its derivatives.

7.1 Aims

Using the aims of the research stated in section 1.6 the following conclusions can be reached.

Aim 1: *To gain an understanding of the CIDOC Conceptual Reference Model (CRM) used to represent the semantic content of cultural data held within museums and art galleries.*

The literature review presented in Chapter 3 of this report identified the key researchers working with CIDOC CRM, primarily Martin Doerr and his associates at FORTH and CIDOC. The review summarised some of the earlier work that appeared instrumental in developing the initial concepts associated with CRM. These included Gruber proposed set of good design criteria subsequently adopted in part by Martin Doerr:

Clarity	Express the meaning of terms in an effective way
Coherence	Definitions should be consistent and meaningful in a natural language.
Extendibility	Flexible and anticipate future changes
Minimal encoding bias	Not dependent on the encoding mechanism
Minimal ontological commitment	Meaning expressed using the minimum of terms.
Tradeoffs	Balance the needs of the above mentioned criteria.

Guarino argued that during the development of a database, the final conceptual model can be displayed as a computer processable ontology, which can be mapped to the

principal target platform. These aspects have been extensively studied for the mapping of the "knowledge specification" to schemas for many different types of database. Guarino claims the utilization of a highly interdisciplinary approach is the main peculiarity of the methodological side of the ontology, and this peculiarity is seen to be one of the most important features of CRM integration for cultural heritage data and information.

An overview and description is provided by Martin Doerr of the CIDOC CRM. He defines CRM as a high level ontology designed to provide definitions and a formal structure, which can be used for describing the implicit and explicit concepts relevant in the area of cultural heritage. The CRM is an ontology formulated in the form of an object-oriented semantic model that aims to solve the problem of semantic interoperability between various kind museum data and their relations to archive and library material.

The review concludes with a description of the CRM structure is described by Doerr and Crofts (1999) together with many of the key terms:

Aim 2: *To develop and use both a graphical and mathematical representation to fully describe the relationships between Entities of the CIDOC CRM.*

In Chapter 4: Analysis 1, the simple graphical format used by Doerr (2000) was significantly modified to cater for the range of new mapping configuration used in validating the CRM. It is believed that the richness embodied in the new graphical format could lead to a better understanding of the CRM.

A meta-taxonomy proposed by Blackwell and Engelhardt was used to analyse the graphical format using nine aspects grouped into four parts: *Signs – components*, *Graphical structure*, *Meaning* and *Context*. The new graphical format suggested the possibility of using notation from category theory. After some basic analysis and making reference to category theorists such as Marquis (2003) it was seen that category theory itself and not just its notation was a real possibility. The formal requirements of category theory and related axioms appear to be satisfied. Some

additional features needed to be added to cater for the extended mapping requirements in the CRM.

Mapping Diagram notation

The following mapping notation was adopted:

- A solid line represents a required morphism
- A interrupted dotted one and only one morphism
- A dotted line represents zero, one or more morphism

Category Theory notation

- Required morphism

$$< p_1.p_2...p_n >: E_o \rightarrow \prod_{1}^{n} E_i$$

- One and only one

$$\left[p_1.p_2...p_n \right]: E_o \rightarrow \overset{n}{\underset{i=1}{\mathbf{C}}} E_i$$

- Zero, one or more

$$[\![p_1, p_2, .. p_n]\!]: E_0 \rightarrow \overset{n}{\underset{i=1}{\square}} E_i$$

The final expression is not available in category theory notation – its use is unlikely to be opposed as it sits well with the previous two situations.

A number of examples were considered that demonstrated that the combination of the mapping diagram and category theory was able to represent the whole range of mapping situations.

Aim 3: *To apply the same graphical and mathematical notation to validate three of the seminal international publications used to validate the CIDOC CRM.*

Three validation activities published in the international literature were replicated using the mapping diagram and category theory.

- EAD – Encoding Archival Description
- DC – Dublin Core
- AMICO – (Art Museum Image Consortium)

These three quite different environments provided the replication exercise with a variety of mapping situations – all of which appear to be appropriately represented by the twin mapping tools. In chapter 5, the researcher follows step by step the mapping

activities of the original authors thus enabling effective comparisons between the different approaches to be made.

The researcher identified two situations where the notation in CRM version 3.0 could be improved.

- Property on Property

  In CRM 3.0, properties such as 'in the role of' do not appear to be attached in any defined way to the 'parent' property, in this case 'p14: carried out by'. The researcher suggests allocating the two properties together. Three mapping situations were identified where this extended notation could apply. As a consequence the 'child' property is given a number.

- Relationship cardinality

  During the mapping exercise it was noted that morphisms could be optional or compulsory. This was not explicitly mentioned in CRM Version 3.0. The researcher made this clear within both the mapping diagram and the mathematical notation.

It is interesting to note that in the most recent version of CRM, Version 3.4.9. (Crofts, et al, 2003) these two shortcomings were addressed by CIDOC. The emphasis on the properties rather than the entities, which was the focus of earlier CIDOC documentation, has been adopted. This is in keeping with this research which, through the use of category theory and the mapping diagram, places more emphasis on the properties.

Aim 4: *To apply the same graphical and mathematical notation to validate the CIDOC CRM using two New Zealand centres of cultural heritage:*

- *Te Manawa – The Palmerston North Science Centre and Museum*
- *The Suter Gallery – Nelson*

Validation exercises were carried out by taking real-life date from records provided by The Suter Gallery in Nelson and Te Manawa in Palmerston North. The types of information recorded by these two sources of cultural heritage were analysed and submitted to the validation process. The data from Te Manawa was taken from a well known Museum Collection Systems (COLLECTIONS – by Vernon Systems Ltd.), and as a result, would enable the findings of this research to be applicable to many art galleries and museums in New Zealand. The Suter Gallery's collection system was

designed and developed in house and it likely to possess a number of unique structures. However, the types of information stored is likely to be typical of small galleries but this has not been confirmed.

Examination of the databases structures and the format of actual records indicated that mapping to the CRM was similar to that of AMICO. The two validaation processes were successful, however, it was thought that the CRM would benefit from adding the extra $E_{xx}Fund$ Entity and Property $p_{xx}fund(is\ funded\ by)$ that are associated with the related 'Loan' information in the museum record.

The four aims, which formed the purpose of this research have been throughly investigated and it would appear that all the objectives have been achieved. It is hoped that the overall intention of the research to add to the understanding of the CIDOC CRM, and how it could support the interoperability of collection systems and the sharing of cultural information, has been achieved.

7.2 Future Research

Further development of the mapping diagram and the category theory notation might provide for an interesting research proposal. There is possibility that the notation could lead to a useful parsing algorithm from database models to the CRM.

There is a possibility too that the CRM and the new notation would lend themselves into representing medical records information or information associated with geographical information systems.

8 References

Art Museum Image Consortium. (2002). *AMICO Date Specification: Data Dictionary Version 1.3*: http://www.amico.org/AMICOlibrary/dataDictionary.html

Art Museum Image Consortium. (2002). *AMICO Date Specification*: http://www.amico.org/AMICOlibrary/dataspec.html

Art Museum Image Consortium. (2002). *AMICO Date Specification: Image& Multimedia. Related Image and Multimedia Files Specification Version 1.2*: http://www.amico.org/AMICOlibrary/dataspec.image.html

Baca, M. and Harpring, P. (Ed.). (2000). *Categories for the Description of Works of Art*. The J. Paul Getty Trust & College Art Association, Inc.: http://www.getty.edu/research/conducting_research/standards/cdwa/index.html

Bertin, J. (1967). *Semiologie graphique: les diagrammes, les reseaux, les cartes*. The Hague, Paris: Mouton and Gauthiers-Villars.

Blackwell, A. F. and Engelhardt, Y. (1998). A taxonomy of diagram taxonomies. In *Proceedings of Thinking with Diagrams 98: Is there a science of diagrams?* (pp.60-70).

Bourner. T. (1996). The research process: four steps to success. In Tony Greenfield (Ed.) *Research Methods: Guidance for Postgraduates*, Arnold, London.

Cox, R. and Brna, P. (1995). *Supporting the use of external representations in problem solving: the need for flexible learning environments.* Journal of Artificial Intelligence in Education, 6(2), pp. 239-302.

Crofts, N., Doerr, M., Gill, T., Stead, S., Stiff, M. (Eds.) (February 2001). *Definition of the CIDOC object-oriented Conceptual Reference Model, Version 3.0,* ICOM/CIDOC Documentation Standards Group. http://cidoc.ics.forth.gr/docs/crm_version_3.0.rtf

Crofts, N., Doerr, M., Gill, T., Stead, S., Stiff, M. (Eds.) (September 2002). *Definition of the CIDOC object-oriented Conceptual Reference Model, Version 3.3.2,* ICOM/CIDOC CRM Special Interest Group. http://cidoc.ics.forth.gr/docs/cidoc_crm_version_3.3.2.rtf

Crofts, N., Doerr, M., Gill, T., Stead, S., Stiff, M. (Eds.). (30th November 2003). *Definition of the CIDOC object-oriented Conceptual Reference Model, Version 3.4.9,* ICOM/CIDOC CRM Special Interest Group.

Dale, E. (1969). *Audiovisual methods in teaching* (3rd edition). New York, Holt, Rhinehart and Winston.

Doerr, M., and Crofts, N. (1999). *Electronic Esperanto: The Role of the Object-Oriented CIDOC Reference Model.* Paper presented at the ICHIM'99, 22-26 September, Washington, D.C.

Doerr, M. (2000). *Mapping of the Dublin Core Metadata Element Set to the CIDOC CRM.* Technical Report FORTH-ICS/TR-274, Foundation for Research and Technology--Hellas, Heraklion, Crete, Greece.

Doerr, M. (2001a). *Mapping of the AMICO Data Dictionary to the CIDOC CRM,* Technical Report 288, ICS-FORTH, Foundation for Research and Technology--Hellas, Heraklion, Crete, Greece.

Doerr, M. (2001b). *The CIDOC CRM – an Ontological Approach to Semantic Interoperability of Metadata.* Foundation for Research and Technology – Hellas, Heraklion, Crete, Greece.

Doerr, M. and Stead, S. (2002): The CIDOC Conceptual Reference Model as a Tool for Integrating Cultural Information. In *Computer Applications and Quantitative Methods in Archaeology Conference*, Heraklion, Crete, Greece.

Doerr, M. (2003, Fall). *The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata*. AI Magazine, Volume 24 Issue 3, pp.75-92.

Dublin Core Metadata Initiative. (2003). *Dublin Core Metadata Element Set, Version 1.1: Reference Description*: http://dublincore.org/documents/dces/

Dublin Core Metadata Initiative. (2003). *DCMI Type Vocabulary*: http://dublincore.org/documents/dcmi-type-vocabulary/

Dublin Core Metadata Initiative. (2000). *Dublin Core Qualifiers*: http://dublincore.org/documents/2000/07/11/dcmes-qualifiers/

Encoded Archival Description Working Group of the Society of American Archivists and the Network Development and MARC Standards Office of the Library of Congress. (2002). *Encoded Archival Description Tag Library (Version 2002)*. Appendix D: Index by Element Name: http://www.loc.gov/ead/tglib/appendix_d.html

Engelhardt, Y. (1998). *Meaningful space: How graphics use space to convey information.* Proceedings Vision Plus 4, School of Design, Carnegie Mellon University, Pittsburgh, pp. 108-126.

Goldsmith, E. (1984). *Research into illustration: An approach and a review.* Cambridge: Cambridge University Press.

Green, T.R.G. and Blackwell, A.F. (1998). *Design for usability using Cognitive Dimensions*. Invited tutorial at BCS HCI'98.

Gruber, T. R. (1993). *Toward Principles for the Design of Ontologies Used for Knowledge Sharing*. Technical Report, KSL93-04, Knowledge Systems Laboratory, Stanford University.

Guarino, N. (1998): Formal Ontology and Information Systems. In Guarino, N (Ed.), *Formal Ontology in Information System. Proceeding of FOIS'98* (pp.3-15). Amsterdam, IOS Press.

Guarino, N. (2003): *Ontological Resources and Top-Level Ontologies*. LADSEB-CNR, Padova, Italy. Retrieved September 2003, from http://www.cs.man.ac.uk/~carole/old/GGF%20Tutorial%20Stuff/Slides%20EU-NSF%20Workshop/swsw-guarino.ppt

Iannella, R. (Ed.). (1999). *DC Agent Qualifiers, DC Working Draft.* Dublin Core Metadata Initiative: http://dublincore.org/usage/decisions/2000/wd-agent-qual.html

ICOM/CIDOC Documentation Standards Group. Retrieved December 2003 from http://cidoc.ics.forth.gr/index.html

Marquis, Jean-Pierre (2003). Category Theory. In Zalta, Edward N. (Ed.) *The Stamford Encyckopedi of Philosophy (Sprin 2003 Edition)*: http://Plato.stanford.edu/archives/spr2003/entries/category-theory/.

Martin, J. and McClure, C. (1985). *Diagramming techniques for analysts and programmers*. Englewood Cliffs, NJ: Prentice-Hall.

Newsham, R. (1995). *Symbolic representation in object-oriented methodologies: Modeling the essence of the computer system*. Unpublished Master's thesis, Department of Computer Science, Nottingham Trent University, United Kingdom.

Payette, S. and Blanchi, C. (May 1999). *Interoperability for Digital Objects and Repositories: The Cornell/CNRI Experiments*, D-Lib Magazine, Volume 5 Issue 5 ISSN 1082-9873; http://www.dlib.org/dlib/may99/payette/05payette.html

Peirce, C.S. (written around 1897, republished in 1932). Elements of Logic. In C. Hartshorne and P. Weiss (Eds.), *The collected papers of C.S. Peirce*. Harvard University Press.

Price, B.A., Baecker, R.M. and Small, I.S. (1993). *A principled taxonomy of software visualization*. Journal of Visual Languages and Computing, 4(3), 211-266.

Saussure, Ferdinand de. (1966) Course in General Linguistics, New York: McGraw-Hill. cited in Peirce, C. (1931-1958). In Hartshorne, C. and Weiss, P. (eds.) *Collected Paper of C.S. Peirce*. Harvard University Press, Cambridge, Massachusetts.

The New Oxford Dictionary (1999). Pesall, J. (Ed.). Oxford University Press.

Theodoridou, M. and Doerr, M. (2001). *Mapping of the Encoded Archival Description DTD Element Set to the CIDOC CRM*. Technical Report 289, ICS-FORTH, Foundation for Research and Technology--Hellas, Heraklion, Crete, Greece.

Vernon Systems Limited (2002). Vernon System Collection. PO Box 6909, Auckland, New Zealand.

9 Appendices

9.1 Appendix 1

Definition of the CIDOobject- oriented Conceptual Reference Model (Version 3.2.1)

9.1.1 The Entity List (Selection)

The following is a subset of the list of all entities and links contained in the model. It consists of an index and the entity declarations themselves. The list is ordered by herarichic level, in a "depth first" manner, from the smaller to the larger subhierarchies, and alphabetically between equal siblings. From this sequence, a unique identifier for each entity emerges, which facilitates cross-referencing.

Entity declarations use the following format:

- Entity names (terms) are presented as headings in bold face, preceded by the unique identifier.
- The line "Belongs to:" refers to the metaclass the entity is a member of.
- The line "Subclass of:" declares the superclass of the entity, from which it inherits links.
- The line "Superclass of:" is a cross-reference to the following subclasses of this entity.
- The line "Scope note" contains the textual definition of the concept the entity represents.
- The title "Properties" announces the list of links.
- Links are grouped by related meaning under metacategories, i.e. a series of titles. e.g. "classifications" etc., in normal face.
- Each link is represented by its forward and backward name, and the entity it links to, separated by colon.
- Links declared directly for the entity are given in bold face.
- Inherited links are given in italics as cross-references to the respective superclasses, for better comprehension.
- Inherited links with a redefined (restricted) target entity are given in bold face italics.
- Each link may be followed by a scope note for the link in an indented text in smaller characters.
- Links of links are given in an indented position in parenthesis under the respective link.

The title "The entity is referenced by:" indicates the cross-reference list of links pointing to this entity (in the sequence called "incoming links"). In cases where there is no such link, the phrase "The entity is not referenced" is used.

Each incoming link is represented by the entity it originates from, and its forward and backward name, separated by a colon, in normal face.

The title "The entity inherits references:" indicates the cross-reference list of links pointing to any of the superclasses of this entity ("inherited incoming links").

Each inherited incoming link is represented by the entity it originates from, and its forward and backward name, separated by a colon, in italics.

9.1.2 E5 Event

Belongs to: Period Type

Subclass of: Period

Superclass of: End of Existence
Beginning of Existence
Activity

Scope note: A change of state in cultural, social, physical systems, regardless of scale, brought about by a series or group of coherent physical, cultural, technological or legal phenomena.

Examples : World War II, Battle of Stalingrad, Earthquake in Lisbon, birth of Cleopatra, my birthday celebration 28-6-1995, the Yal ta Conference, "a tile fell from my roof", the CIDOC Conference 2005.

The distinction between and event and a period is partly a question of scale. Viewed at a broad scale, an event is an 'instantaneous' change of state. At a fine scale, the event can be analysed into its component phenomena within a space and time frame, i.e., a period. The reverse is not necessarily the case, not all periods give rise to a noteworthy change of state.

9.1.2.1 Properties

identifications

is identified by (identifies): Appellation

classifications

has type (is type of): Type

active participants

had participants (participated in):Actor

property note: this is the superproperty of "carried out by", "has formed", "by mother", "brought into life", "dissolved", "was death".

passive participants

occurred in the presence of (was present at): Stuff

property note: this is the superproperty of "destroyed", "used object", "transferred title of", "moved", "transferred custody of", "has modified", "concerned", "registered", "measured", "has created".

spatial definitions

took place at (witnessed): Place

spatial definitions, short cut

took place on or within (witnessed): Physical Object

temporal definitions

has time-span (is time-span of): Time-Span

structures

consists of (forms part of): Period

falls within (contains): Period

other descriptions

has note: String

(has type : Type)

9.1.2.2 The entity is referenced by:

Physical Man-Made Stuff: depicts event (is depicted by)

(mode of depiction : Type)

9.1.2.3 The entity inherits references:

Period: consists of (forms part of)

Period: falls within (contains)

Type Assignment: classified (was classified by)

Document: documents (is documented in)

Conceptual Object: refers to (is referred to by)

(has type : Type)

9.1.3 E19 Physical Object

Belongs to: Physical Object Type

Subclass of: Physical Stuff

Superclass of: Biological Object

Man-Made Object

Scope note: A discrete, real item of material nature which constitutes a unit for documentation. The decision as to what constitutes a complete item, rather than parts or components, may be purely administrative.

Examples : John Smith, Aphrodite of Milos, the Palace of Knossos, the Cullinan diamond, Apollo 13 a the time of launch.

9.1.3.1 Properties:

identifications

is identified by (identifies): Object Identifier

property note: this is the subproperty of "E1CRM Entity.(is identified by)", and superproperty of "E19 Physical Object.(preferred identifier is)".

preferred identifier is (is preferred identifier of): Object Identifier

property note: this is the subproperty of "E19Physical Object.(is identified by)".

is identified by (identifies): Appellation

classifications

has type (is type of): Type

legal status

is subject to (applies to): Right

legal status, short cut

has former or current keeper (is former or current keeper of) : Actor

property note: this is the superproperty of "has current keeper". It is short cut of the path "Custody changed by – Transfer of Custody – custody received / surrendered by".

has current keeper (is current keeper of) : Actor

property note: this is the subproperty of "has former or current keeper". It is short cut of the path "Custody changed by – Transfer of Custody – custody received / surrendered by".

has former or current owner (is former or current owner of): Actor

property note: this is the superproperty of "has current owner". It is short cut of the path "Acquisition changed by – Acquisition – transferred title from / to"

has current owner (is current owner of): Actor

property note: this is the subproperty of "has former or current owner". ". It is short cut of the path "Acquisition changed by – Acquisition – transferred title from / to"

right held by (has right on): Actor

(has type: Type)

(has note: String)

physical status, short cut

has dimension (is dimension of): Dimension

has condition (condition of): Condition State

locations, short cut

has former or current location (is former or current location of) : Place

property note: this is the superproperty of "has current permanent location", "has current location". It is short cut of the path "Move changed by – Move – moved from / to"

has current permanent location (is current permanent location of): Place

property note: this is the subproperty of "has former or current location". It is short cut of the path "Move changed by – Move – moved from / to"

has current location (currently holds) : Place

property note: this is the subproperty of "has former or current location". It is short cut of the path "Move changed by – Move – moved from / to"

structures

bears feature (is found on): Physical Feature

has number of parts: Number

has section definition (defines section): Section Definition

is composed of (forms part of): Physical Stuff

consists of (is incorporated in): Material

structures, short cut

has section (is located on or within): Place

property note: It is short cut of the "Section Definition"

other descriptions

had as general use (was use of): Type

has note: String

(has type : Type)

9.1.3.2 *The entity is referenced by:*

Period: took place on or within (witnessed)

Destruction: destroyed (was destroyed by)

Activity: used object (was used for)

(mode of use: String)

Acquisition: transferred title of (changed ownership by)

Move: moved (moved by)

Transfer of Custody: transferred custody of (custody changed by)

Identifier Assignment: registered (was registered by)

9.1.3.3 *The entity inherits references:*

Physical Man-Made Stuff: depicts object (is depicted by)

(mode of depiction : Type)

Physical Stuff: is composed of (forms part of)

Condition Assessment: concerned (was assessed by)

Measurement: measured (was measured by)

Event: occurred in the presence of (was present at)

Type Assignment: classified (was classified by)

Document: documents (is documented in)

Conceptual Object: refers to (is referred to by)

(has type : Type)

Beginning of Existence: brought into existence (was brought into existence by)

End of Existence: took out of existence (was taken out of existence by)

9.1.4 E7 Activity

Belongs to:	Period Type
Subclass of:	Event
Superclass of:	Formation
	Conceptual Creation
	Modification
	Transfer of Custody
	Acquisition
	Move
	Attribute Assignment

Scope note: An action or a series of actions, carried out by actors (people, groups or organisations) which follow a certain explicit or implicit intention and result as a collective effect in some change of state in the cultural, social, physical systems we are interested in. This notion includes both complex and long lasting actions such as the building of a settlement, or a war, as well as simple, short-lived actions such as the opening of a door. It does not include the notion of activity in the sense of professions and other non-targeted notions. These are seen rather as belonging to a part in the hierarchy above Event.

9.1.4.1 Properties

identifications

is identified by (identifies): Appellation

classifications

has type (is type of): Type

active participants

carried out by (performed): Actor

(in the role of : Type)

property note: this is the superproperty of "transferred title to", "transferred title from", "custody surrendered by", "custody received by" and subproperty of "had participants".

had participants (participated in): Actor

passive participants

took into account (was taken into account by): Conceptual Object

property note: The equivalent of using something physical. May be better "used: Stuff".

used object (was used for): Physical Object

(mode of use: String)

property note: this is the subproperty of "occurred in the presence of".

occurred in the presence of (was present at):Stuff

motivations

was motivation for (motivated): Conceptual Object

motivated the creation of (was created for): Conceptual Object

was intended use of (was made for): Man-Made Stuff

(mode of use: String)

had specific purpose (was purpose of): Activity

had as general purpose (was purpose of): Type

spatial definitions

took place at (witnessed): Place

spatial definitions, short cut

took place on or within (witnessed): Physical Object

temporal definitions

has time-span (is time-span of): Time-Span

structures

consists of (forms part of): Period

falls within (contains): Period

other descriptions

has note: String

(has type : Type)

9.1.4.2 The entity is only referenced by itself

9.1.4.3 The entity inherits references:

Physical Man-Made Stuff: depicts event (is depicted by)

(mode of depiction : Type)

Period: consists of (forms part of)

Period: falls within (contains)

Type Assignment: classified (was classified by)

Document: documents (is documented in)

Conceptual Object: refers to (is referred to by)

(has type : Type)

9.2 Appendix 2

Index of the properties of the CIDOC CRM sorted by domain

Property id	Entity – Domain	Domain id	Property Name	Entity - Range	Range id
P1	CRM Entity	E1	is identified by (identifies)	Appellation	E41
P2	CRM Entity	E1	has type (is type of)	Type	E55
P3	CRM Entity	E1	has note (has type : Type)	String	E62
P4	Temporal Entity	E2	has time-span (is time-span of)	Time-Span	E52
P5	Condition State	E3	consists of (forms part of)	Condition State	E3
P6	Condition State	E3	falls within (contains)	Condition State	E3
P7	Period	E4	took place at (witnessed)	Place	E53
P8	Period	E4	took place on or within (witnessed)	Physical Object	E19
P9	Period	E4	consists of (forms part of)	Period	E4
P10	Period	E4	falls within (contains)	Period	E4
P11	Event	E5	had participants (participated in)	Actor	E39
P12	Event	E5	occurred in the presence of (was present at)	Stuff	E70
P13	Destruction	E6	destroyed (was destroyed by)	Physical Object	E19
P14	Activity	E7	carried out by (performed) (in the role of : Type)	Actor	E39
P15	Activity	E7	took into account (was taken into account by)	Conceptual Object	E28
P16	Activity	E7	used object (was used for) (mode of use : String)	Physical Object	E19
P17	Activity	E7	was motivation for (motivated)	Conceptual Object	E28
P18	Activity	E7	motivated the creation of (was created for)	Conceptual Object	E28
P19	Activity	E7	was intended use of (was made for) (mode of use: String)	Man-Made Stuff	E71
P20	Activity	E7	had specific purpose (was purpose of)	Activity	E7
P21	Activity	E7	had as general purpose (was purpose of)	Type	E55
P22	Acquisition	E8	*transferred title to (acquired title of)*	*Actor*	E39
P23	Acquisition	E8	*transferred title from (surrendered title of)*	*Actor*	E39
P24	Acquisition	E8	Transferred title of (changed ownership by)	Physical Object	E19
P25	Move	E9	moved (moved by)	Physical Object	E19
P26	Move	E9	*moved to (occupied)*	*Place*	E53
P27	Move	E9	*moved from (vacated)*	*Place*	E53
P28	Transfer of Custody	E10	*custody surrendered by (surrendered custody)*	*Actor*	E39
P29	Transfer of Custody	E10	*custody received by (received custody)*	*Actor*	E39
P30	Transfer of Custody	E10	Transferred custody of (custody changed by)	Physical Object	E19
P31	Modification	E11	has modified (was modified by)	Physical Man-Made Stuff	E24
P32	Modification	E11	used general technique (was technique of)	Type	E55
P33	Modification	E11	used specific technique (was used by)	Design or Procedure	E29
P108	Production	E12	has produced (was produced by)	Physical Man-Made Stuff	E24
P34	Condition Assessment	E14	concerned (was assessed by)	Physical Stuff	E18
P35	Condition Assessment	E14	has identified (identified by)	Condition State	E3
P36	Identifier Assignment	E15	registered (was registered by)	Physical Object	E19
P37	Identifier Assignment	E15	assigns (is assigned by)	Object Identifier	E42
P38	Identifier Assignment	E15	deassigns (is deassigned by)	Object Identifier	E42

Property id	Entity – Domain	Domain id	Property Name	Entity - Range	Range id
P39	Measurement	E16	measured (was measured by)	Physical Stuff	E18
P40	Measurement	E16	observed dimension (was observed)	Dimension	E54
P41	Type Assignment	E17	classified (was classified by)	CRM Entity	E1
P42	Type Assignment	E17	assigned (was assigned by)	Type	E55
P43	Physical Stuff	E18	has dimension (is dimension of)	Dimension	E54
P44	Physical Stuff	E18	has condition (condition of)	Condition State	E3
P45	Physical Stuff	E18	consists of (is incorporated in)	Material	E57
P46	Physical Stuff	E18	is composed of (forms part of)	Physical Stuff	E18
P47	Physical Object	E19	is identified by (identifies)	Object Identifier	E42
P48	Physical Object	E19	preferred identifier is (is preferred identifier of)	Object Identifier	E42
P49	Physical Object	E19	has former or current keeper (is former or current keeper of)	Actor	E39
P50	Physical Object	E19	has current keeper (is current keeper of)	Actor	E39
P51	Physical Object	E19	has former or current owner (is former or current owner of)	Actor	E39
P52	Physical Object	E19	has current owner (is current owner of)	Actor	E39
P53	Physical Object	E19	has former or current location (is former or current location of)	Place	E53
P54	Physical Object	E19	has current permanent location (is current permanent location of)	Place	E53
P55	Physical Object	E19	has current location (currently holds)	Place	E53
P56	Physical Object	E19	bears feature (is found on)	Physical Feature	E26
P57	Physical Object	E19	has number of parts	Number	E60
P58	Physical Object	E19	has section definition (defines section)	Section Definition	E46
P59	Physical Object	E19	has section (is located on or within)	Place	E53
P60	Person	E21	is member of (has members)	Legal Body	E40
P61	Person	E21	has gender (is gender of)	Gender	E76
P62	Physical Man-Made Stuff	E24	depicts object (is depicted by) (mode of depiction : Type)	Physical Stuff	E18
P63	Physical Man-Made Stuff	E24	depicts event (is depicted by) (mode of depiction : Type)	Event	E5
P64	Physical Man-Made Stuff	E24	depicts concept (is depicted by) (mode of depiction : Type)	Type	E55
P65	Physical Man-Made Stuff	E24	shows visual item (is shown by)	Visual Item	E36
P66	Conceptual Object	E28	refers to concept (is referred to by)	Type	E55
P67	Conceptual Object	E28	refers to (is referred to by) (has type : Type)	CRM Entity	E1
P68	Design or Procedure	E29	usually employs (is usually employed by)	Material	E57
P69	Design or Procedure	E29	is associated with	Design or Procedure	E29
P70	Document	E31	documents (is documented in)	CRM Entity	E1
P71	Authority Document	E32	contains (is part of)	Type	E55
P72	Linguistic Object	E33	has language (is language of)	Language	E56
P73	Linguistic Object	E33	has translation (is translation of)	Linguistic Object	E33
P74	Actor	E39	has current or former residence (is current or former residence of)	Place	E53
P75	Actor	E39	possesses (is possessed by)	Right	E30
P76	Actor	E39	has contact points (provides access to)	Contact Point	E51
P77	Legal Body	E40	consists of (belongs to)	Legal Body	E40
P78	Time-Span	E52	is identified by (identifies)	Time Appellation	E49
P79	Time-Span	E52	begins at qualify	String	E62
P80	Time-Span	E52	ends at qualify	String	E62
P81	Time-Span	E52	at least covering	Time Primitive	E61
P82	Time-Span	E52	at most within	Time Primitive	E61
P83	Time-Span	E52	had at least duration	Dimension	E54
P84	Time-Span	E52	had at most duration	Dimension	E54
P85	Time-Span	E52	consists of (forms part of)	Time-Span	E52
P86	Time-Span	E52	falls within (contains)	Time-Span	E52
P87	Place	E53	is identified by (identifies)	Place Appellation	E44
P88	Place	E53	consists of (forms part of)	Place	E53
P89	Place	E53	falls within (contains)	Place	E53
P90	Dimension	E54	value	Number	E60

Property id	Entity – Domain	Domain id	Property Name	Entity - Range	Range id
P91	Dimension	E54	unit	Measurement Unit	E58
P92	Beginning of Existence	E63	brought into existence (was brought into existence by)	Existence	E77
P93	End of Existence	E64	took out of existence (was taken out of existence by)	Existence	E77
P94	Conceptual Creation	E65	*has created (was created by)*	Conceptual Object	E28
P95	Formation	E66	*has formed (was formed by)*	Group	E74
P96	Birth	E67	by mother (gave birth)	Person	E21
P97	Birth	E67	from father (was father for)	Person	E21
P98	Birth	E67	*brought into life (was born)*	Person	E21
P99	Dissolution	E68	*dissolved (was dissolved by)*	Group	E74
P100	Death	E69	*was death of (died in)*	Person	E21
P101	Stuff	E70	had as general use (was use of)	Type	E55
P102	Man-Made Stuff	E71	has title (is title of) (has type : Type)	Title	E35
P103	Man-Made Stuff	E71	was intended for (was intention of)	Type	E55
P104	Legal Object	E72	is subject to (applies to)	Right	E30
P105	Legal Object	E72	right held by (has right on) (has type : Type) (has note : String)	Actor	E39
P106	Information Object	E73	is composed of (forms part of)	Information Object	E73
P107	Group	E74	had member (was member of)	Actor	E39

9.3 Appendix 3

AMICO Data Specification: Data Dictionary Version 1.3 (2002)

Record Type	TAG	AMICO-FIELD	Min. Cite?	Core?	Repeat?	Group?	Definition/Guidelines	Examples, separated by end of field delimiter }~
Catalog Record Fields (One Record for each Work of Art)								
Unique Identification								
Catalog	AID	AMICO Identifier	*	*	N		A unique identifier, assigned to a work in the AMICO Library. Comprised of a 4 letter institutional abbreviation, followed by a dot, and then a unique number such as an accession number [DOIs will be investigated in the future]	NMAA.87-32547/a-g}~ AIC_456502}~ GEH_3457-86}~
What is it?								
Catalog	OTY	Object-Type		*	N		The kind of work of art described; chosen from a short list of terms	installation}~ sculpture}~ watercolor}~
Catalog	OPP	Object-Parts/Pieces			Y		The number and a description of any parts/pieces of the work of art	chair, shelf, painting on canvas, robe, and bricks}~ 3 panels}~ 2}~
Catalog	CLG	Classification			Y	Group	Does not contain data; used to group fields classifying work.	}~
Catalog	CLT	Classification-Term			N	CLG	Terms used to associate this work with other like works	sculptural multimedia installation}~ works of art on paper}~
Catalog	CLS	Classification-Scheme			N	CLG	The classification scheme from which a term was	AAT}~
What is it called?								
Catalog	OTG	Object-Title/Name	*		Y	Group	Group element	}~
Catalog	OTN	Object-Title-Name		*	N	OTG	The title or name of the work	In the Afternoon}~ untitled}~ Blue #6}~
Catalog	OTT	Title-Type			Y	OTG	The kind of Title or Name assigned to the work: Can include terms/phrases such as 'preferred', 'as given by artist' etc.	preferred}~ popular}~ as first exhibited}~
Catalog	OST	State			Y		For works produced in multiples, the state of this particular impression	1 of 5}~ only known}~ artists proof}~
Catalog	OEN	Edition			Y		For works produced in multiples, the edition of this particular example.	1st}~ Second American}~
What does it look Like?								
Catalog	OPD	Physical Description			N		A narrative description of the physical appearance of the work including any parts or components.	The objects are off center with a brick pathway leading towards them and the canvas in the background}~ Installation piece for the 2nd floor sculpture court comprised of several pine sticks. The ones in the center are burned.}~
Catalog	OPA	Physical Orientation/ Arrangement			N		A narrative description of the orientation of the work, or its physical arrangement or set-up. For example, used to describe how to assemble an installation, or hang an oddly shaped or abstract work	horizontal}~ designed for 14'x28' room}~ on pedestal}~
Catalog	MET	Measurements-Text		*	Y		A free text display form of the works measurements.	14'x28'}~ 17.6 " x 38" unframed}~ 160 kg.}~ 13.2 cu. m.}~
Catalog	MEG	Measurements			Y	Group	Does not contain data; used to group fields recording measurements.	}~
Catalog	MCM	Measurement-Component-Measured			N	MEG	A term indicating what was measured (sheet, frame, lid, pedestal etc.)	sheet}~ frame}~ lid}~ pedestal}~
Catalog	MED	Measurement-Dimension			N	MEG	A term indicating the measurement taken, drawn from the following list: height, width, depth, weight, circumference, duration, volume (other terms may be added)	height}~ width}~ length}~ circumference}~ diameter}~ volume}~ weight}~ duration}~
Catalog	MDV	Measurement-Dimension-Value			N	MEG	The number of units of the measurement (numeric only)	37.6}~ 14}~ 182.25}~
Catalog	MDU	Measurement-Dimension-Units			N	MEG	The unit in which the measurement was taken: inches, feet, mm, cm, lbs., oz., kg, gr., minutes, sq. ft., cu. ft. etc.	linear measurements: inches}~ ft}~ mm.}~ cm}~ meters}~ planar measurements: sq.ft.}~ sq m.}~ volumetric measurements: cu.ft. cu. meters}~ pounds, ounces}~ grams, kilograms}~ seconds, minutes}~ hours}~

Record Type	TAG	AMICO-FIELD	Min. Cite ?	Core?	Repeat ?	Group?	Definition/Guidelines	Examples, separated by end of field delimiter }~
Catalog Record Fields (One Record for each Work of Art)								
What does it look Like?								
Catalog	MEQ	Measurement-Qualifier			N	MEG	A qualification or indication of the accuracy of a measurement	plus/minus 10}~ accurate to 96%}~ circa}~
Catalog	OMG	Materials and Techniques			Y	Group	Does not contain data; used to group fields documenting materials and techniques used to create the work	}~
Catalog	OMD	Materials and Techniques-Description		*	N	OMG	A free text description of the materials and techniques used to create the work of	Bricks, furniture and canvas}~ ink and watercolor on paper}~ oil on canvas}~
Catalog	OMT	Materials and Techniques-Process/Technique-Term			Y	OMG	Single terms that index the processes and techniques used to create the work.	construction}~ freehand drawing}~ drip painting}~ lithograph}~
Catalog	OMM	Materials and Techniques-Materials-Term			Y	OMG	Single terms that index the materials used to create the work.	brick}~ oak}~ watercolor}~ varnish}~
Catalog	OMS	Materials and Techniques-Support			Y	OMG	Single terms that index the support on which the work was created.	rice paper}~ canvas}~
Catalog	OIN	Inscriptions and/or Marks			Y		A free text description or transcription of any inscriptions or marks on the work, including their location, medium, hand, and other details	signed, JG., lower front}~ silver mark on handle}~ inscribed "to my friend john" on rear, with date July 14, 1973}~
Catalog	OCH	Condition/Examination History			Y		A narrative description of the Condition or Examination history of the work of art. Cite documentation associated with Condition or Examination History in the Related Documents Group	Excellent Condition according to Condition Report, 14/7/82, H.M. Black, Conservator}~ not inspected when first accessioned, found to have scratch in upper left, 3 inches long, during 1965 inventory}~
Catalog	OTH	Treatment/Conservation History			Y		A narrative description of the Treatment or Conservation of the Work of art. Cite Documentation associated with Treatment or Conservation in the Related Documents Group	Restored, summer 1987. Detailed treatment report available}~ Base reaffixed July 1987}~
Who made it?								
Catalog	CRG	Creator			Y	Group	Does not contain data; used to group the fields documenting the creator of thew work.	}~
Catalog	CRQ	Creator-Qualifier			N	CRG	A qualification of the attribution of the work to a particular creator.	School of}~ Follower of}~ Attributed to}~ Copy after}~
Catalog	CRT	Creator-Name-Text	*	*	N	CRG	Display form (direct order) of Creator Name (and qualifier). If Creator-Name isn't present, display form of Creator Culture Nationality	Vito Acconci}~ W. Eugene Smith }~ Rembrandt van Rijn}~
Catalog	CRN	Creator-Name		*or CRC	N	CRG	Sort Form (inverted, last name first) of Creator name; used in indexing.	Acconci, Vito}~ Smith, W. Eugene}~ Rembrandt van Rijn}~
Catalog	CRC	Creator-Culture/Nationality		* or CRN	N	CRG	The culture or nationality of the creator who made the work, or to which the creation of the work is attributed	Benin Culture}~ Egyptian 1085-710 B.C.}~ Native American}~
Catalog	CDT	Creator-Dates/Locations-Text			N	CRG	A free text description of dates and places associated with the creator or culture that created the work.	born 1876 in Staffordshire, England}~ after 800 B.C.}~ not before 1766}~ c.1450}~ 3rd century B.C.}~ Died 1990}~
Catalog	CBD	Creator-Birth-Date			N	CRG	The date of birth of the creator.	1957}~ 19780914}~ 190206}~
Catalog	CBP	Creator-Birth-Place			N	CRG	The place the creator was born.	Lille, France}~ The Netherlands}~ London, Ontario, Canada}~
Catalog	CBQ	Creator-Birth-Qualifier			N	CRG	A free text qualifier of the date of birth of the creator.	c.}~ circa}~ before}~ not after}~ no later than}~
Catalog	CDD	Creator-Death-Date			N	CRG	The date the creator of the work died.	1320}~ 19380914}~ 710 B.C.}~
Catalog	CDP	Creator-Death-Place			N	CRG	The place of death of the creator.	Helsinki, Finland}~ India}~ at sea, South Atlantic}~
Catalog	CDQ	Creator-Death-Qualifier			N	CRG	A free text qualifier of the date of death of the creator.	c.}~ circa}~ before}~ not after}~ no later than}~
Catalog	CAD	Creator-Active-Date			N	CRG	The creator's dates of activity; may be known when birth and death are not.	1342-1386}~ 1960's}~ early 7th century B.C.}~
Catalog	CAP	Creator-Active-Place			Y	CRG	The place where the creator was active. May repeat with multiple places.	France}~ Middle Kingdom, Egypt}~ North America}~
Catalog	CGN	Creator-Gender			N	CRG	the gender of the creator.	Male}~ Female}~
Catalog	CRB	Creator-Biography			N	CRG	A free text biography of the creator of the work.	Active in Canada 1844-71; Came to Canada in 1842}~ Winner of the Prix du Rome in 1903.
Catalog	CRR	Creator-Role			Y	CRG	A term or terms describing the role played by the creator in the making of the work of	sculptor}~ designer}~ print maker}~ castings manager}~
Catalog	CNO	Creator-Notes			N	CRG	A free text note about the creator, and the relationship between the creator or culture and the work of art.	Not known to have ever signed works.}~
Catalog	CID	Creator Identification Number				CRG	The identification number for the creator as assigned in the AMICO Artists Reference File	AMICO: 20001

Record Type	TAG	AMICO-FIELD	Min. Cite?	Core?	Repeat?	Group?	Definition/Guidelines	Examples, separated by end of field delimiter }~
Catalog Record Fields (One Record for each Work of Art)								
When was it made?								
Catalog	OCG	Creation-Dates			Y	Group	Does not contain data; used to group fields dating work's creation.	}~
Catalog	OCT	Creation-Date-Text	*	*	N	OCG	The date on which a work was created, or a range of dates during which it could have been created	c. 1645}~ 1957}~ dated by the artist as September 13, 1975}~
Catalog	OCS	Creation-Date-Start			N	OCG	A number, indicating the earliest possible date a work could have been created; BC dates are recorded as negative integers	1300}~ 198209}~ 19461103}~
Catalog	OCE	Creation-Date-End			N	OCG	A number, indicating the latest possible date a work could have been created; BC dates are recorded as negative integers	1350}~ 198301}~ 19461103}~
Catalog	OCQ	Creation-Date-Qualifier			N	OCG	A qualifier, that indicates approximation in the earliest or latest date.	c.}~ circa}~ before}~ not after}~ no later than}~
Where was it made?								
Catalog	OCP	Creation-Place			Y		A place or places where the work was created.	Paris, France}~ Loire River Valley}~ Los Angeles, California, USA}~
What is it about?								
Catalog	STG	Style/Period			Y	Group	Does not contain data; used to group fields describing work's style or period.	}~
Catalog	STD	Style/Period-Description			N	STG	A narrative description of the style or period of the work of art.	Early Christian with Byzantine influence}~ Eastern Han ceramic ware with 13th century decorative glazes}~ Baroque}~ Etruscan influenced}~
Catalog	STT	Style/Period-Terms			Y	STG	Index terms that characterize the style and/or period of the work of art.	Art Nouveau}~ Minyan ware}~ French Colonial}~
Catalog	SUG	Subject Matter			Y	Group	Does not contain data; used to group fields documenting work's subject matter.	}~
Catalog	SUP	Subject Matter-PreIconographic Description			N	SUG	A free text description of the generic subject of the work of art.	Pastoral landscape with cattle in fields in background}~ Full length portrait with ball gown and pearls}~ Group of men at table gambling and drinking}~
Catalog	SUI	Subject Matter-Iconography			Y	SUG	A free text description of the specific, named subject of the work of art	Diana and hounds}~ Christ in the temple}~ Herrod and the slaughter of the innocents}~
Catalog	SUT	Subject Matter-Index Terms			Y	SUG	Index terms that characterize the subject of the work of art.	Mrs. John Patorius}~ Court of Louis XIV}~ Netherlands - 16th century, Recreation}~
Catalog	CXG	Context			Y	Group	Does not contain data; used to group fields documenting work's context.	}~
Catalog	CXD	Context-Description			N	CXG	A narrative description of the historical context of the work of art, including its creation, display, excavation, or other history.	Excavated at Pompeii in the 1876 expedition}~ Installed in Brussels town square until 1782}~ Originally displayed with framelike structure built by the artist as part of the Black3 series}~
Catalog	CXP	Context-Related-Person			Y	CXG	Index forms of the names of any people contextually related to the work of art.	Pershing, John}~ Emperor Hirohito}~ Wu Hen Din}~ Pope Julius II}~
Catalog	CXS	Context-Related Site/Place			Y	CXG	The names of any places that are contextually related to the work of art.	Brussels, Belgium}~ Pompeii, Italy}~ Times Square, New York, New York, USA}~
Catalog	CXT	Context-Time Period/Dates			N	CXG	The dates, times or periods of a particular context.	1876}~ 1782}~ prior to acquisition by the museum in 1956}~}~
What does it mean?								
Catalog	OCR	Critical Responses			Y		Narrative discussions of the critical reception or analysis of the work of art. Link any critical texts in the Related Documents Group.	"Delightful! - A triumph." Maravius, Hans, "In the Light of day", The New City News, January 7, 1987}~
Who showed it ?								
Catalog	OEH	Exhibition or Loan History			Y		A record of when and where the work has been exhibited.	Manchester Museum of Art, June 10 - September 18, 1976, "New Lines in Grand Masters"}~
Catalog	OOG	Owner			Y	Group	Does not contain data; used to group fields documenting works ownership.	}~

Record Type	TAG	AMICO-FIELD	Min. Cite ?	Core?	Repeat ?	Group?	Definition/Guidelines	Examples, separated by end of field delimiter }~
Catalog Record Fields (One Record for each Work of Art)								
Who owned it?								
Catalog	OOG	Owner			Y	Group	Does not contain data; used to group fields documenting works ownership.	}~
Catalog	OON	Owner Name	*	*	N	OOG	The name of the institution or individual who owns the work now [e.g. the AMICO member]. Can repeat if multiple/part owners	SF MOMA}~ The Trustees of the Harold J. Jameson Memorial Museum of Crafts}~
Catalog	OOP	Owner-Place		*	Y	OOG	The place where the work is owned.	Buffalo, New York, USA}~ St. Petersburg, Russia}~ San Diego, Chile}~
Catalog	OOA	Owner-Accession-Number		*	Y	OOG	The accession or inventory number assigned to the work by this owner.	87-4532.a-c}~ 456565-0112}~
Catalog	OOC	Owner-Credit-Line	*	*	Y	OOG	The Credit Line, or any required mention or acknowledgement of the ownership of the work of art.	Gift of Mrs. John Francis Blue}~ Purchased with funds from trustees and friends in memory of Hector Escobosa, Brayton Wilbur and J.D. Zellerbach}~
Catalog	OPO	Provenance/Prior Owners-Text			Y		A record of the past owners of the work of art.	Andrew Carnegie}~ ASEA International Ltd.}~ Mr. & Mrs. John A. T. Pastors}~
Catalog	ORG	Rights/Copyright			Y	Group	Does not contain data; used to group fields documenting work's copyright or restrictions.	}~
Catalog	ORS	Copyright-Statement	*		N	ORG	The copyright of the work of art, including any known rights holders, encumbrances or restrictions.	Permission for educational use only granted by the Trustees of the Picasso estate}~ No reproductions permitted under terms of the gift}~
Catalog	ORL	Copyright-Link		*	N	ORG	A URL that links to the AMICO member, indicating where and how a user can acquire further information or rights to use the work.	http://www.artic.edu}~
What is it related to?								
Catalog	RWG	Related Works of Art			Y	Group	Does not contain data; groups fields documenting related works of art.	}~
Catalog	RWD	Related-Works-Description			N	RWG	A narrative description of the relationship between this work and any others.	David Hockney, "Mulholland Drive: the Road to the Studio", 1980}~ Other statues from the same site accessioned by the British Museum in 1845}~
Catalog	RWR	Related-Works-Relationship-Type			N	RWG	The kind of relationship; drawn from the list of Dublin Core Relationship types.	IsBasisFor}~ IsPartOf}~ References}~
Catalog	RWL	Related-Works-Identifier/Link			N	RWG	An identifier for or link to the related work of art.	LAMA.AC1954.856.32}~ AKAG.1863:1}~
Catalog	RIG	Related Images		*	Y	Group	Does not contain data; groups fields documenting related images. All AMICO works must have at least one related image.	}~
Catalog	RIP	Related-Image-Preferred		*	N	RIG	Indicates [yes/no] whether this is the preferred image of the work of art. Each work must have one preferred image, that will be used in lists and brief displays.	Y}~ N}~
Catalog	RID	Related-Image-Description		*	N	RIG	The view of the work shown in this image: full, detail, recto, verso, 360 degrees, etc.; Note: Maps to XDE in Multimedia Metadata Record	Detail}~ Full View}~ Aerial view}~ 360 degrees}~x-ray}~
Catalog	RIR	Related-Image-Relationship-Type		*	N	RIG	The relation between this work and the image, using values from the Dublin Core Relationship Types. If original work is analog, and the image is digital, this value will be "HasFormat". If the original work is digital and the image is digital, this value	HasFormat}~ HasVersion}~
Catalog	RIL	Related-Image-Identifier/Link		*	N	RIG	The identifier of the related image' Note: maps to XID in Multimedia Metadata Record	AIC_.96-34543.tif}~ SFMA.89-2335.jpg}~
Catalog	RMG	Related Multimedia			Y	Group	Does not contain data; groups fields documenting related multimedia files.	}~
Catalog	RMD	Related-Multimedia-Description			N	RMG	A description of the related multimedia file. Note: Maps to XDE in Multimedia Metadata Record	Interview, in the artist's studio, January 17, 1993}~ "Fly-Through" movie of the installation seen from three feet above ground}~ Interpretation offered by the Museum Director in the gallery audio}~
Catalog	RMR	Related-Multimedia-Relationship-Type			N	RMG	The relationship between this work and the related multimedia file, expressed using Dublin Core Relationship Types. Note: inverse in XRT of Multimedia Metadata Record	References}~ IsVersionOf}~ IsBasedOn}~
Catalog	RML	Related-Multimedia-Identifier/Link			N	RMG	The identifier of the related multimedia file. Note: maps to XID in Multimedia Metadata Record	SFMA.96-543.mov}~ SDCA.1324:87.1.wav}~ NMAA.67-3452.cgm}~
Catalog	RDG	Related Documents			Y	Group	Does not contain data; groups fields describing related documents.	}~
Catalog	RDD	Related-Document-Description			N	RDG	The description, or title with full citation, of the related document. Note: Maps to XDE in Multimedia Metadata Record	Johnson, James, "Major Works by Hungarian Artists, London, Ballet & Baney, 1963 p.137-138}~ The New Reporter, "ARTopenARTclosed", June 17, 1989, p.7}~ transcript of gallery talk at opening, Museum Archives}~
Catalog	RDR	Related-Document-Relationship-Type			N	RDG	A description of the relationship between the work and the related document, expressed using Dublin Core Relationship Types. Note: inverse of XRT in Multimedia Metadata Record	References}~
Catalog	RDL	Related-Document-Identifier/Link			N	RDG	The Identifier of the Related Document. Note: maps to XID in Multimedia Metadata	WGA_.doc1876-45.sgm}~ MIA_.newreporter46.gif}~ NMAA.5x85-342.txt}~

Record Type	TAG	AMICO-FIELD	Min. Cite ?	Core?	Repeat ?	Group?	Definition/Guidelines	Examples, separated by end of field delimiter }~
Catalog Record Fields (One Record for each Work of Art)								
Who documented it?								
Catalog	DCG	Documentation/ Cataloguing-History			Y	Group	Does not contain data; groups fields recording the documentation history of the work.	}~
Catalog	DCB	Documented/Cataloged By			N	DCG	The name of the person who documented the work.	Jameson}~ PDT}~ PhotoServices3}~
Catalog	DCD	Documented/Cataloged-Date			N	DCG	The date the work was documented.	19870415}~
Media Metadata Fields (One record for each related image or media file)								
Media Metadata	XID	DC.Resource.Identifier		*	N		The identifier of the metadata file. Maps to RIL, RML or RDL in catalog record. Validation ensures conformity to file naming conventions	MMA_.39504.TIF AIC_.MUM98r.MOV
Media Metadata	XTI	DC. Title			Y		The title or name of the work in the media file; maps to RID, RMD, or RDD in the catalog record.	Front view}~
Media Metadata	XCN	DC.Creator			Y	Group	Does not contain data; groups fields documenting the creation of the media file.	}~
Media Metadata	XCP	DC.Creator.PersonalName			Y	XCN	The Personal Name of the creator of media file	
Media Metadata	XCC	DC.Creator.CorporateName			Y	XCN	The Corporate Name of the Creator of the media file.	
Media Metadata	XCR	DC.Creator.Role			Y	XCN	The role the named creator played in the making of the media file.	
Media Metadata	XDE	DC.Description		*	Y		A narrative description of the contents of the media file.	
Media Metadata	XPU	DC.Publisher		*	N		The name of the institution that made the media file available; I.e. the name of the contributing AMICO member.	Art Institute of Chicago}~
Media Metadata	XDN	DC.Contributor			Y	Group	Does not contain data; groups fields documenting other contributions to the creation of the media file.	
Media Metadata	XDP	DC.Contributor.PersonalName			Y	XDN	The Personal Name of a contributor to the creation of the media file	
Media Metadata	XDC	DC.Contributor.CorporateName			Y	XDN	The Corporate Name of a contributor to the creation of the media file.	
Media Metadata	XDR	DC.Contributor.Role			Y	XDN	The role played by the contributor in the creation of the media file.	
Media Metadata	XDA	DC.Date			Y		The date that the media file was created, in the format YYYYMMDD	
Media Metadata	XRT	DC.ResourceType		*	N		A "genre-like" term, characterizing the content of the media file. For example, analytical report; critical review; interview; letter; lecture; portrait; reproduction;	reproduction}~
Media Metadata	XAM	AMICO.Mode		*	N		A term indicating the modality in which a person will experience the resource. audio; image; model; multimedia; text; video	image}~
Media Metadata	XFO	DC.Format			Y		Not Used: AMICO records format in specific sub-elements.	
Media Metadata	XFE	AMICO.Format.Encoding		*	N	XFO	The manner in which the data in the media file is encoded. E.g. gif; tiff; jfif; ascii	TIFF}~
Media Metadata	XFP	AMICO.Format.ColorPalette			N	XFO	The color palette of the media file. E.g. RGB, CMYK, Grayscale	RGB}~

Record Type	TAG	AMICO-FIELD	Min. Cite?	Core?	Repeat?	Group?	Definition/Guidelines	Examples, separated by end of field delimiter }~
Catalog Record Fields (One Record for each Work of Art)								
Media Metadata Fields (One record for each related image or media file)								
Media Metadata	XCM	AMICO.Format.ColorMetric			N	XFO	The manner in which the color data is represented; e.g., 24-bit, 8-bit.	24-bit}~
Media Metadata	XFD	AMICO.Format.Dimensions		*	N	XFO	The dimensions of the media file: For an image, the spatial resolution of the image, e.g. 1024 x 768; 640 x 480; etc.	1024 x 768}~
Media Metadata	XFF	AMICO.Format.FileSize		*	N	XFO	The storage size of the media file.	2.7 MB}~
Media Metadata	XFC	AMICO.Format.Compression		*	N	XFO	The compression algorithm used in storing the media file: none; jpeg; mpeg; etc.	JPEG}~ MPEG}~ QT}~
Media Metadata	XRE	DC.Relation			Y	Group	Does not contain data; used to group fields documenting relationships.	
Media Metadata	XRY	DC.Relation.Type		*	N	XRE	The kind of relationship, expressed as a Dublin Core Relationship type. Note: this is the inverse of the relation in the record pointed to by the Relation Identifier	IsFormatOf}~ IsVersionOf}~
Media Metadata	XRI	DC.Relation.Identifier		*	N	XRE	The identifier of the related work, image, metadata or sound; If within the AMICO Library, this must conform to naming conventions; warning produced if points outside AMICO dataset	AIC_.253846v}~
Media Metadata	XRS	DC.Rights	*	*	Y		Text of rights statement for this media file.	Copyright Art Institute of Chicago}~
Media Metadata	XMN	AMICO.Media.Note			N		Capture details and other matters of importance for understanding the quality/content of the digital file	
Media Metadata	XVD	AMICO Metadata Validation Date		*	N		Added by AMICO: Date the file was validated	19980506}~
Media Metadata	XVV	AMICO Data Dictionary Version		*	N		Version of the Data Dictionary	1
Media Metadata	XPR	Metadata Data Processing Note		*	Y		Added by AMICO: Documents data processing routines, errors and additions through parsing routines	WARNING: XRI does not point to a valid AID}~
Media Metadata	XDL	Metadata Delition Flag			N		Entered by AMICO. Indicates that a record should be removed from all public distributions of the library. Y=deleted record	Y}~
Media Metadata	XLY	Metadata Library Year			N		Entered by AMICO. Indicates which library year this record was originally received in.	1998}~

9.4 Appendix 4

Correspondence relating to The Suter Gallery.

All the communication with the Gallery was via email and over the telephone although the initial contact was made in February 2003 at a presentation given by this researcher in Nelson.

Sent: Thursday, 18 March 2004 3:59 p.m.
To: Zhou, Jia
Subject: Re: Assistance

From: Zhou, Jia
To: The Suter - Collection
Sent: Thursday, March 18, 2004 2:08 PM
Subject: RE: Assistance

Hello Marie,

It's really nice to get your feedback regards to the fields description, and I believe the document is going to help me a lot with my research. However the message I got from you is without the attachment, could you please send the attachment once more time?

Many thanks,

Jia
-----Original Message-----
From: The Suter - Collection [redacted]
Sent: Monday, 15 March 2004 11:37 a.m.
To: Zhou, Jia
Subject: Re: Assistance

Hello Jia,

Thank you for your enquiry.

I've attached a Word document of The Suter data records fields for you.

I hope this helps.

Marie Claude
Collection Technician
The Suter
The Aratoi o Whakatu

----- Original Message -----
From: Zhou, Jia
To: The Suter - Collection
Sent: Tuesday, March 02, 2004 3:01 PM
Subject: RE: Assistance

Hello Judith,

It's Jia Zhou from Massey University, thanks for your help on supporting me get further understanding of your data record fields last time.

I'm at the stage of mapping the Suter data to the semantic modelling I'm working on, both my supervisor (Barry Jackson) and I think that it will be much nicer if we would have a full description of all the Suter data record fields from you. It will be great for me if you have the relevant information.

Thanks and best regards,

Jia
-----Original Message-----
From: The Suter - Collection [redacted]z]
Sent: Saturday, 10 January 2004 12:58 p.m.
To: Zhou, Jia
Subject: Re: Assistance

Hi Jia

I have put the answers to your questions next to your text below- dont hesitate to ask more questions if you need to.
Best wishes for the New Year to you and Barry.
Judith

Accession No - how does this differ from the Suter Record number? The Suter record number is a number for the record only. ie the record holds information relating to a particular accession number. The accession number is the number that is designated to the work itself , its recorded physically on the work, is the main way of tracking the object and remains with it for its life in the institution. It is often used to identify a work before title or artists name (which can sometimes be confusing eg if an artist has more than one work with the same or similar title for instance).

Support - this is the surface or material that holds or supports the artwork eg for an oil (the medium) on canvas the canvas is the support. For a watercolour (the medium) on paper-the paper is the support.

Support Auxiliary - this is where there is more than one support. There can be a secondary support eg sometimes paper is glued to cardboard. So a watercolour can be on paper on cardboard. The cardboard would be the auxiliary support.

Construction- this is related to how the work is made. Especially where a work is multimedia. It helps us to look after a work and display it by understanding how a work is made. For example, a .work might be constructed from fabric and paper glued with a particular glue and attached to a frame of some sort. Construction knowledge helps us to know how fragile a work might be, or how to best store it, the kind of conditions it might need for preservation,transportation etc.

Obj Disposal - at times in the past some artworks have been disposed of -this might have been because they were thought to be damaged beyond repair and no further

use existed for them.This rarely happens anymore.

Picture - refers to any known illustration of a work in a publication

Notes - this is any additional contextual information or notes mostly about ongoing research about an object. I have used it when people have offered me unsubstantiated information about an object without references. It could serve as a place for anything that doesn't fit another category.

Inventory - we carry out routine physical inventories of the whole collection. It is updated by physical sighting of a work and record the date the object was last seen on.

Photo: Medium/Index - we have a numbering system for photographs/negatives/transparencies of collection items.

I hope this helps.

Judith

The following is a copy of the attachment.

COLLECTION DATABASE

FIELD NAME	DESCRIPTION
Number	Auto number
Accession No	Number from the original Accession register. This must be a number. No letters or other characters may be used.
Work Title	Name of work as given by artist.
Object	Describes what object is: Sketch, Drawing, Ceramic etc.
Glaze	This field refers to ceramic works only. Describes what glaze is used and glaze colours.
Decoration	This field refers to ceramic works only. Type of decoration used if any.
Date created	Date work was finished.
Description	Contains descriptive material as necessary.
Inscriptions	Marks or signature put on work by artist. Also includes Potters marks on Pottery works
No of Pieces	Indicates the number of pieces in the complete work.
Weight	Contains weight of object particularly three-dimensional works.
Location	Records location
Medium	Indicates material used to produce work; Oil, Acrylic, Crayon etc.
Support	Describes what the work was produced on; Canvas, Paper, Board etc. Use one indicator only i.e. the primary support.
Support auxiliary	Describes material paper or canvas may be attached to i.e. paper (glued on card) should be described by one term only in this case paper (in support) and then Card in additional support. Additional materials (eg Glue) should be part of construction
Construction	Describes how the object is constructed.
Dim Height mm	Dimension Height in millimetres.
Dim Width mm	Dimension Width in millimetres.
Dim Depth mm	Dimension Depth in millimetres.
Dim Diameter mm	Dimension Diameter in millimetres.
Provenance	Describes how the works first came to be in the Suter Collection; Donated, Bequeathed etc.
Donor Name	Records original donors name.
Artwork Exhibitions	Records any exhibitions the work has been used in. Should give the name of the exhibition, date and place.
Loans	Records any loans of works. Includes dates, reasons, and places.
Artist	Contains artists name and initials. No punctuation to be used.
Courtesy Title	Sir, Lady etc.
Forenames	Forenames where known.
Honours	Records any honours, which may have been bestowed on artist.
Date of Birth	Records Year of Birth.
Date of Death	Records Year of Death.
Nationality	Records nationality or place of birth.
Biographical Details	This field contains a brief biography of the artist. See artist table. Also see artist table.
Chronology	Contains career info in chronological order and may stand in place of biographical details.
Obj Constraints	Records any publications, which may refer to the artist.
Artist Exhibitions	Records any exhibitions the artist has had works in.
Obj Condition	Records general notes on condition of work.
Obj Conservation Report	Contains detailed notes on damage, drying, cracking etc.
Obj Treatment	Contains detailed notes on treatment methods required to restore work to original condition.
Obj Sponsor	Records names and date of conservation sponsors
Obj Constraints	Records any constraints, which may apply to the exhibition of any work.
Obj Copyright	Records any Copyright conditions, which may apply to reproduction of work.

Obj Mode of Accession	Records method of Accession; A=Allocated, L=Loan, D=Donated, Presented or Gifted, B=Bequeathed, P=Purchased
Obj Date of Accession	Records original accession date. Data must be entered as per input mask.
Obj Deaccession	Yes means the work is no longer is the collection. 215 works listed in 2002
Obj Deaccession Date	Records date work is deaccessioned or removed from collection.
Obj Disposal	Records method of disposal.
References	Contains reference to any published material about the work.
Photo References	Contains photo references where available.
Picture	Contains graphical reproduction of work for reference purposes.
Notes	This field is a catchall for any information, which does not readily fit into any other field.
Cataloguer/Data Entry	Name of data entry person.
Date Catalogued	Date of last entry.
Photo: Photographer	Name of Photographer
Photo: Medium/Index	CP=Colour Print C N=Colour Negative Strip CT=Colour Transparency CS=Colour Slide BWP=Black and White Print D=digital filed on CD
Inventory	Date work sighted for inventory purposes. Verified physical check work in collection. Items not inventoried are assumed missing, not located or deaccessioned.
Original Purchase price	Price paid at time of purchase
Not located	Item missing from collection, no note, date or explanation for deaccession available.
Valuation for insurance	Valuation based on market valuation of similar works sold within the last 4 years. Value is set of replacement value in current market and remains current for one year from May 2002 i.e. 1/3 higher than ordinary retail value.

9.5 Appendix 5

Correspondence relating to Te Manawa.

The collection management system used by Te Manawa was developed and installed by Vernon Systems Ltd. The company was extremely helpful in providing information to support this research but was concerned that some aspects of their data schema should not be published for commercial reasons. A non-disclosure form was signed.

From: Abby Turbott, Vernon Systems Ltd[redacted]
Sent: Friday, 12 March 2004 11:28 a.m.
To: Zhou, Jia
Subject: Re: Assistance

Hi Carol,

Thanks - we just got your faxed non-disclosure form. Attached is a list
of Vernon Cataloguing Object fields and their suggested use.

Good luck with your Masters thesis, and do let us know if we can be any further assistance.

Regards, Abby

--
Abby Turbott
Systems Consultant

Vernon Systems Ltd.
[redacted]

Zhou, Jia wrote:

Dear Bil,

It's Carol from Massey University, I had contact with Abby Turbott
earlier this year about my master thesis on researching an approach to help support Internet interoperability between centres of cultural heritage. The main thrust of the research is a type of semantic model. which permits mapping of cultural data stored in local databases to this common

model. The proposed model *CIDOC' has been developed and has been 'validated' by its developers with EAD, AMICO, SPECTRUM, Dublin Core and several other museum and art gallery collections. You may wish to visit their site. The International Committee for Documentation of the International Council of Museums (ICOM-CIDOC):
http://www.willpowerinfo.myby.co.uk/cidoc/.

To validate the model, I chosen The Te Manawa Museum to be one of the mapping practice examples. I visited Te Manawa Museum two months ago, and I got some of the printout of their category records. Now I'm at the stage of mapping some of the relative fields to the CRM model. To get a further understanding of those fields, I would like to get a copy of the description of the fields. Here is an example of the description of the field "AMICO Identifier" I got from the AMICO on-line Library: "AMICO Identifier: A unique identifier, assigned to a work in the AMICO Library. Comprised of a 4 letter institutional abbreviation, followed by a dot."

However, I got a message from Susanne in Te Manawa (see the message from Susanne below) that I may need your help to get these definitions, since they don't have any. I'm just wondering if I can get the support from you on this. I'm attaching those fields(attributes) I collected from Te Manawa, hope you can help me on the definition of these fields.

It's will be so great to get the support from you again, and I am sorry to be taking time from your work.

Best wishes,

Jia (Carol)

A meeting with staff at Te Manawa resulted in a number of documents being provided. None of these are include in this thesis ina ccordance with the wishes of Vernon Systems Limited.

-----Original Message-----
From: Susanne Geiser [[redacted]z]
Sent: Tuesday, 13 January 2004 3:40 p.m.
To: Zhou, Jia
Subject: database printouts

Dear Jia,

Cindy and I have done a few database printouts for you.
Can you please email me the address you want me to send them to?

Please be aware that some of the information, esp. re: lenders and valuations is highly confidential. Also, our database system is fairly new, which is why some information is rather basic or brief. Eventually Te Manawa ART will add digital images for all of the works onto the database, a project I am doing later in the year.

Please let me know if we can be of further help to you.

Susanne Geiser
Registrar (Art & Touring Exhibitions)

Te Manawa
Museum I Gallery I Science Centre
LIFE I ART I MIND

396 Main Street
Private Bag 11055
Palmerston North
New Zealand

T 64 6 355 5000 ext. 7078
F 64 6 358 8849
[redacted]z
W http://www.temanawa.co.nz