The Representation of Archival Descriptions: An Ontological Approach

Qing Zou

School of Information Studies, McGill University

Montreal, Quebec, Canada

A thesis submitted to McGill University in partial fulfillment of the requirements
of the degree of Doctor of Philosophy

DEDICATION

Dedicated to my wife Lanfen, for her endless love and patience,

to my boys Alexandar and Ethan,

to my parents, brother, and sister whose unconditional love is the source of inspiration throughout challenging times.

谨以此文，感谢我的妻子, 儿子(可爱的一凡和一树),

以及生我养我的父母大人,

一贯支持我的弟弟和妹妹。

ACKNOWLEDGEMENT

This dissertation would never have been possible with the support of many people. I would like to express my sincere gratitude to all the people who helped me finish this dissertation.

Firstly, the deepest gratitude goes to my advisor, Dr. Eun Gyung Park, for the continuous support of this study, for her patience and immense knowledge, and for guiding me throughout my long dissertation journey. Her guidance helped me in all the time of research and writing this dissertation. She constantly encouraged me to think critically and work diligently. I am so grateful to have her as my advisor for my study.

Besides my advisor, I would like to thank the members of my thesis advisory committee: Dr. Charles-Antoine Julien and Dr. Benjamin C. M. Fung for their insightful comments and valuable recommendations to widen my research and improve my dissertation.

I would thank Mr. Gordon Burr who reviewed my dissertation and provided valuable comments that helped me improve my dissertation.

I am also grateful to Dr. Elaine Ménard as an advisory committee member who provided comments to the early draft of this dissertation.

Lastly, I would like to thank the participants of the focus groups who kindly agreed to take part in my study.

ABSTRACT

Archival descriptions play multiple roles in understanding and managing records in archives. However, archival descriptions are not efficiently organized and presented on the Web. In particular, archival context is often lacking in current descriptions, which makes it difficult for users to search and access records. To address the challenges of current archival descriptions in web environments, the overarching question is asked: *How can archival context be improved for searching, using and understanding in web environments*? In addition, it is necessary to build a conceptual/operational model for representing and restructuring the archival context and provenance information of archival descriptions. Two research questions are formulated to address the two objectives of the study, which are 1) to identify and examine the essential components for archival context and their relationships in archival descriptions, and 2) to explore whether archival context can be explicitly represented and modeled using ontology technologies.

To answer the research questions, a multi-method is taken in three studies: 1) Study 1 is a content analysis of archival descriptions; 2) Study 2 is a methontology to develop an event-based ontology; and 3) Study 3 is a focus group study to evaluate the usefulness of the built ontology. In Study 1, sample archival descriptions are drawn from 40 archival institution sites. By conducting the content analysis, the ten essential entities from archival descriptions have been identified as *Agent*, *Artifact*, *Event*, *Feature*, *Function*, *Place*, *Relation*, *Role*, *Situation*, and *Time*. In Study 2, the results of Study 1 were used to build the ontology model. By taking a methontology approach, five stages are performed: to identify ontology requirements at the specification stage, to identify relationships and concepts at the conceptualization stage, to develop the ontology at the implementation stage, to align with other ontologies based on the Basic Formal Ontology at the formalization stage, and to validate the ontology at the evaluation stage. As a result, an Event

Ontology for Archival Context (EOAC) is developed with 233 classes and 87 properties, which were explicitly formalized in OWL (Web Ontology Language) for implementation. In Study 3, two focus group sessions with users and professional archivists were conducted to elicit both users' and professional opinions on the current status of archival descriptions and the ontology model. Participants explained the advantages and disadvantages of the current descriptions and the ontology model and ten out of eleven participants were willing to adopt the event ontology model if supporting tools were available.

The findings of the three studies led to the development of a comprehensive conceptual model for archival context. The model consists of three levels: (1) the micro-level with static and fluid views; (2) the meso-level with three separate contexts (i.e., creation, description, and usage contexts); and (3) the macro-level with a series of activity systems. By integrating all three levels, the EOAC provides an explicit and comprehensive way to transform archival context into more structured, connected and interactive forms of entities.

This research adopts an innovative event-centred approach and emphasizes that events are as important as objects. The three-level conceptual model for archival context provides a framework to deepen the understanding of archival context. The EOAC ontology provides a new way to transform archival contexts into Linked Open Data, which can be utilized to improve searches and retrievals of archival descriptions and meet the expectations of users.

RÉSUMÉ

Les descriptions archivistiques jouent plusieurs rôles importants dans la compréhension et la gestion des registres d'archives. Cependant, sur la toile, l'organisation et la présentation de ces descriptions ne sont pas efficaces. L'absence de contexte archivistique dans les descriptions actuellement en ligne rend la recherche et l'accès à ces documents particulièrement difficile pour les usagers. Afin d'aborder les défis que présente cette situation, la question qui suit guidera l'ensemble de la recherche : *Comment le contexte archivistique peut-il être amélioré sur la toile pour des fins de recherche, d'utilisation et de compréhension?* De plus, il est nécessaire de construire un modèle conceptuel opérationnel qui représentera et restructurera le contexte archivistique et la provenance de l'information des descriptions archivistiques. Les objectifs de cette étude sont donc les suivants : 1) identifier et examiner les composantes essentielles du contexte archivistique ainsi que leur relation aux descriptions archivistiques et 2) définir si les technologies de l'ontologie peuvent être utilisées pour représenter le contexte archivistique de manière plus explicite. Deux questions de recherche sont alors explorées : 1) quelles sont les composantes essentielles du contexte archivistique et 2) dans quelle mesure les technologies de l'ontologie peuvent-elles être utilisées afin de représenter le contexte archivistique?

Afin de répondre aux questions de recherche, une approche méthodologique mixte servira lors de trois études : 1) la première étude est une analyse de contenu des descriptions archivistiques, 2) la deuxième étude utilise la méthontologie comme méthodologie pour développer une ontologie basée sur les événements et 3) la troisième étude évalue l'utilité de l'ontologie développée par le moyen de groupes de discussion. Dans la première étude, des échantillons de descriptions archivistiques sont recueillis sur les sites de quarante institutions. En effectuant l'analyse de contenu, dix composantes ont pu être identifiées comme étant essentielles aux descriptions

archivistiques : exécutant, artefact, événement, caractéristique, fonction, lieu, relation, rôle, situation et temps. Dans la deuxième étude, les résultats obtenus lors de la première étude sont utilisés en suivant la méthontologie pour construire le modèle ontologique. Le modèle se développe en cinq étapes : identification des besoins de l'ontologie à l'état de spécifications, identification des relations et des concepts à l'état de conceptualisation, développement de l'ontologie à l'étape d'implémentation, alignement avec d'autres ontologies basées sur le *Basic Formal Ontology* à l'état d'élaboration et, finalement, validation de l'ontologie à l'étape d'évaluation. Le résultat est le développement d'une ontologie basée sur les événements, nommée *Event Ontology for Archival Context (EOAC)*, et contenant 233 classes et 87 propriétés, formalisées explicitement dans le language OWL (*Web Ontology Language)* pour l'implémentation. Dans la troisième étude, deux groupes de discussion ont été dirigés auprès d'utilisateurs et d'archivistes professionnels afin de connaître les opinions des utilisateurs et des professionnels sur les descriptions archivistiques actuelles ainsi que sur le nouveau modèle ontologique. Les participants ont pu identifier les avantages et les désavantages des deux méthodes et dix participants sur onze étaient prêts à utiliser la nouvelle ontologie développée si les outils nécessaires étaient mis à leur disposition.

Les conclusions de ces trois études permettent donc le développement d'un modèle conceptuel complet du contexte archivistique à trois niveaux : (1) le micro-niveau, comportant des vues statiques et fluides, (2) le méso-niveau, comportant trois contextes distincts (création, description et contexte d'usage), et (3) le macro-niveau, comportant plusieurs systèmes d'activités. En intégrant les trois niveaux, le modèle EOAC permet de transformer de manière explicite et compréhensive le contexte archivistique en une forme plus structurée, connectée et interactive.

L’approche centrée sur les événements utilisée dans cette recherche est innovatrice et met l’accent sur le fait que les événements sont tout aussi importants que les objets. L’approche conceptuelle à trois niveaux fourni une base qui permet d’approfondir la compréhension du contexte archivistique. L’ontologie EOAC propose une nouvelle façon de transformer le contexte archivistique en données ouvertes liées, ce qui améliorera les recherches et l’extraction des descriptions archivistiques et répondra aux attendes des utilisateurs.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

1.1. Statement of the problem

An archival description is a set of information that faithfully describes what an archival record is about (*content*) in archives. This includes the following information: when, where, and how the record was created and used; who created and used it; in which social and environmental surroundings (*context*) it was used; how the record is organized in a fonds or collection; the order of files in a record; whether the record has been further grouped into small units (*structure*), etc. Thus, archival description is the process of establishing administrative and intellectual control over archival holdings. It is considered the core product of appraisal and arrangement resulting from archival processes. The critical piece of information in an archival description is provenance, which indicates "the origin or source of something" and contains "information regarding the origins, custody, and ownership of an item of a collection" (Pearce-Moses, 2005, p. 317). Provenance is the crucial part of archival context as well as the foundation of archives.

As archival descriptions contain a great amount of information, they also play multiple roles in understanding and managing records in archives. In particular, archival descriptions work as surrogates between users and archives by providing gateways for users to access the archives. According to the Rules for Archival Description (Bureau of Canadian Archivists, Planning Committee on Descriptive Standards, 2008), archival descriptions consist of several elements, such as title, physical characteristics (e.g., the number of boxes of files or cubic meters of files), creator(s), dates, biographical information of the creator(s) (in cases with records about a person or family), an administrative summary, content and scope, and the way in which files are organized (e.g., whether files are arranged by subject or chronologically, etc.). The degree of details on the information included in the archival description can have an impact on how users find the archival record due to the uniqueness of archival records. This aspect implies that an archival description is difficult and complicated but indispensable. For example, one cubic meter of image files can be described by only one archival description record. If one piece of information is not recorded in the archival description, users have no way of knowing whether files with such information even exists or not. Archivists rely on archival descriptions to lead users to archives.

To faithfully describe archival records, archival descriptions should include the content, context, and structure of records as the fundamental components for describing a fonds/collection. More than that, a question on the amount and scope of detailed information that are included in the content, context, and structure is still significant to archivists when creating archival descriptions. For example, the Yellowknife Journal (Steinbruck, 1999), which is a birch bark journal by a North West Company fur trader named Jean Steinbruck, provides an idea of what should be included in the context information. This journal is a short but valuable source that vividly presents the activities of a fur trader from 1802 to 1803 and his interactions with the environment and relationship with the Aboriginal people. However, with regards to the provenance of this journal, the journal poses several challenges to archivists. Steinbruck and the North West Company are undoubtedly a part of the provenance, but in his journal, when Steinbruck quoted Aboriginals, he operated in an isolated post, used birch bark instead of paper, wrote the journal in French, and mentioned his Aboriginal wife and their children. With current practices, fairly obvious information, such as the title of the creator, dates, and functions, was recorded with little exploration of historical, social, cultural, and technological contexts of that time.

Another example is the Madeleine Parent fonds, taken from the McGill University Archives, using a typical biographical sketch, scope, and content. An except from the archival description of Madeleine Parent states, [She] "received her early education … in English at the

Trafalgar School for Girls" and "then attended McGill University, graduating with a B.A. in 1940"…. "[r]etiring from union work in 1983 to Montreal, Madeleine Parent continued her social-activist role, focusing on women's rights. She became a founding member of the National Action Committee on Status of Women (NAC)…" (McGill University Archives, 2018). People who read the archival description need to know it is rare that girls attend schools and graduate from universities at that time, in order to better understand the social status of Madeleine Parent's family and the achievement of her graduation from McGill University. In addition, the Trafalgar School is a prestigious school for girls and McGill University is also a prestigious university. Furthermore, the description does not mention Léa Roback, who convinced Madeleine Parent to join the union movement. The description may provide an impression that her involvement in NAC was after 1983, as NAC was founded in 1971 and she was the representative for Quebec between 1972 and 1983 (Sabourin, 2017).

These examples demonstrate that contextual descriptions do not sufficiently cover the societal, cultural, and technological aspects of archival records. Nesmith (2006) remarked that context in archival descriptions is "still mainly viewed at its surface level" (p. 352). The problem of describing context in the current archival description practices should address broad social dimensions because context information related to societal, cultural and technological dimensions is often lacking in archival descriptions (Millar, 2002). Due to insufficient/confusing contextual information in the current archival descriptions, when users browse and search for information from archival descriptions on the Web, it is difficult for them to find proper archival descriptions and understand what is described in the archival context.

Furthermore, although the importance of archival descriptions has been highly recognized in archival communities, another challenge is that archival descriptions are often kept in a text format in documents or relational databases (e.g., MySQL/MariaDB databases are used in archival management systems such as AtoM and ArchivesSpace). In Web environments, archival description documents are commonly presented in HTML (Hyper Text Markup Language), Microsoft Word format, or PDF (Portable Document Format) files, or XML (eXtensible Markup Language) formats. For example, in the presidential libraries in the United States, 1,074 archival descriptions from the Clinton Digital Library are stored in PDF files (Clinton Digital Library, 2018). Records are described in the Dublin Core metadata format (e.g., title, description, identifier, and publisher) and organized in Omeka, which is a digital asset management system. Detailed archival descriptions consist of a summary, scope and content, and system of arrangement that appear in long paragraphs kept in PDF files. Only two elements (e.g., title, publisher) are searchable in the Clinton Digital Library. In addition, the archival descriptions of the George W. Bush Presidential Library and Museum (The George W. Bush Presidential Library and Museum, 2018) are also stored and listed in PDF files only without any searchable information. Even in AtoM or ArchivesSpace, archival descriptions are stored in relational databases with searchable interfaces and archival descriptions can be exported as EAD (Encoded Archival Description) XML documents. This alleviates the burden to deal with EAD manually but still falls int o the category of a traditional web of documents. For example, the above-mentioned Madeleine Parent fonds was partially moved to AtoM for now so that the fonds level descriptions could be viewed as a web page or EAD file. However, the series and file lists of the Madeleine Parent fonds are still stored in a PDF file. Even with a completed fonds in AtoM the biographical sketch of the fonds is presented in large chunks of text that cannot be directly utilized by search engines or software applications.

As can be observed by these examples of two presidential libraries and AtoM managed archival sites, archival descriptions are stored in PDF/EAD files at many libraries, archives, and museums, which makes it difficult for users to navigate related archival records from one library to another. On the Web, hyperlinks make it possible to link from one piece of information to another. For information producers, although it may save time to present documents in PDF format, PDF files are not able to link documents within and across collections on the Web. Moreover, search engines need to parse PDF files in order to extract the content of the files (e.g., the extracted content may be different from the original content of the files, and some of the PDF files are not extractable). When users need extra software programs to view PDF files, it is difficult for them to navigate PDF documents across collections. This non-searchable PDF document brings obstacles to users when they are searching and using archival descriptions on the Web. Even with EAD documents, they can be presented as HTML pages with long lists that are not effective or user-friendly (Toms, 2001). With regards to these problems, Nimer and Daines (2008) pointed out that archival descriptions are not explicitly and efficiently organized and presented on the Web. In this way, EAD archival descriptions can be used with halfway technology only (Dow, 2009).

The World Wide Web was initially developed as a web of documents, where "digital documents link to each other directly, allowing the user to follow the pointers provided by the author from a place in one document to another digital document" (Coyle, 2010, p. 12). Typical examples of digital documents include HTML web pages and PDF files. With recent emerging technologies, a smart (action) web can support tasks such as searching for items automatically through smart applications, which enables a web of documents to be transformed to a web of data (Berners-Lee, Hendler, & Lassila, 2001). This Semantic Web vision was further enhanced in 2006 with a Linked Data approach that outlines four rules to transform a web of documents into a web of data with Semantic Web technologies (Berners-Lee, 2006). The linked data provides the Web with an "environment where applications can query that data, draw inferences using vocabularies" (W3C, 2015, para. 1). This shift has a great impact on modeling and presenting all types of information on the Web in the fields of information science, libraries, and museums. Among Semantic Web technologies, ontologies refer to a "formal, explicit specification of a shared conceptualization" (Studer, Benjamins, & Fensel, 1998, p. 184) and can offer advanced capabilities to deal with certain types of semantics in information organization. As part of the semantic technologies, ontology is especially suitable for data modeling on the Web.

In archival communities, archival descriptions still seem to be produced in the traditional format: a web of documents even in archives by using archival management systems (e.g. AtoM or ArchivesSpace). As Yoose & Perkins (2013) noted libraries, archives, and museums have accumulated "an embarrassment of riches in the form of unique digitized resources and structured data as well as unmined, unstructured content, all of which are lying fallow inside a Web of documents and untapped relationships" (p. 209). There have been relatively few discussions on the organization and representation of archival descriptions by using Semantic Web technologies in the field of archives compared to other relevant areas. Since archival descriptions on the Web of documents inherit the problem of the Web, the linked data movement is gradually becoming popular, such as addressed by Gracy (2015) and the Linked Jazz project (Pattuelli, Weller, & Szablya, 2011). There have been attempts to convert EAD files into a linked data form (e.g., Archives Hubs). However, since the creation of the EAD in the late 1990s, to convert EAD files to a linked data form, what should be included in archival descriptions has not yet been clearly defined and solved in archival communities. As the importance of archival context has been widely emphasized in the field, characteristics of archival context have been discussed, including archival context is indefinite (Duff & Harris, 2002), complex (Duchein, 1983; Hurley, 2005; Millar, 2002; Niu, 2013), and in multiple aspects (Foscarini, 2010; Lemieux, 2014; Niu, 2013). Some researchers and archivists (Nesmith, 2006; Piggott, 2012) are aware of the societal, cultural, and technological dimensions of archival context. Regardless, to date, there has been no explicitly defined framework or model to examine archival context in depth. The challenges that are stated at the beginning of this section (e.g., the current archival descriptions are not sufficient, difficult to search on the Web, context information should be included in archival descriptions, etc.) are symptoms of a deeply rooted problem: there is a lack of a conceptual/operational model for organizing archival descriptions, particularly provenance and contextual information, regardless of the essential parts of archival descriptions. Therefore, this research is designed to fill this gap. More specifically, this research focuses on the archival context of archival descriptions and aims to address the overarching question*: How can archival context be improved for searching, using and understanding in web environments*?

To address the identified challenges, this research proposes investigating and scrutinizing the elements of current archival descriptions, identifying the underlying relationships of elements and exploring a new way to model the essential elements of archival context through the adoption of ontology technologies. The new model will be developed to better organize and present archival context as an integral part of archival theory and practice.

1.2. Objectives of the study

This study aims to examine archival context in a systematic way and reshape archival context in the form of ontology models. More specifically, the study has two objectives:

1) To identify and examine the essential components for archival context and their relationships in archival descriptions; and

2) To explore whether archival context can be explicitly represented and modeled using ontology technologies.

Based on analyses of archival descriptions and description representations in ontologies, a new model will be designed to describe archival context.

1.3. Research questions

As archival descriptions are one of the most important processes and products in archives, it is critical to identify the fundamental components of archival descriptions included in archival context. To better understand and formally express archival context, and to achieve the objectives of this study, research questions are further divided into the following two sets of sub-questions.

Question 1: What entities are essential components for archival context?

1.1 What entities are used to provide information on archival context in existing archival descriptions?

1.2 What are the relationships of those entities in representing archival context?

Question 2: To what extent can ontology technologies be used for representing archival context?

2.1 What components are necessary for representing archival context when using ontology technologies? How can ontology technologies be used to represent information on archival context?

2.2 Are ontology technologies perceived to be appropriate for representing archival context?

Question 1 aims to identify the essential components of archival context, which is the curial part of archival descriptions and will be examined to identify the relationships among the identified entities. Question 2 aims to explore the application of ontology technologies to archival context and examine their applicability to users and professionals.

This dissertation is organized into seven chapters and appendices. In Chapter 1, the research problem statements and research questions are presented. In Chapter 2, the two major foundations of the subject of this study (e.g., archival descriptions and the Semantic Web and ontology) are explained. The methodologies of this dissertation are described in Chapter 3. The findings of the study are presented in Chapters 4 and 5. The built model and important issues drawn from the findings are discussed in Chapter 6. The conclusions, contributions, and future directions of the study are presented in Chapter 7.

1.4. Significance of the study

This study primarily emphasizes the importance of archival context and examines archival context. It provides a comprehensive conceptual model for analyzing archival context, which helps contribute to archival theory and practice.

Most of all, this study applies ontology technologies to build an event ontology for archival context. This study attempts to adopt an event-centred approach instead of a traditional object-centred approach. This research on event-centred archival descriptions will help contribute to advancement of archival descriptions and open a new way of organizing and improving archival descriptions. The findings of the study will help provide a path or model towards the use of Linked Open Data and promote the shift from a web of documents to a web of data in archival science and other relevant fields.

1.5. Definition of terms

It is important to define the key terms used in this research to clarify the study. Key terms are selected from two constructs: archives and the Semantic Web. The definitions given below are not strict but are used to help understand the research. Formal definitions and detailed explanations are provided in Chapter 2. The definitions are as follows:

Activity: "A series of acts or actions aimed at one purpose" (The InterPARES project team, 2018, para. 1).

Agent: "Something that bears some form of responsibility for an activity taking place, for the existence of an entity, or for another agent's activity" (Lebo, Sahoo, & McGuinness, 2013, sec. 4.1).

Archival description: The process of establishing physical, administrative, and intellectual control over archival holdings and the outcomes of such a process (Pearce-Moses, 2005).

Archives: Recorded information regardless of stored medium with historical, informational, or evidential value produced in transactions or activities (Pearce-Moses, 2005).

Attribute: "A feature or characteristic; a property" (Pearce-Moses, 2005, p. 39).

Class: A group of things sharing common characteristics (Pearce-Moses, 2005, p.72).

Collection: "A group of materials with some unifying characteristic, assembled by a person, organization, or repository from a variety of sources" (Pearce-Moses, 2005, p.76). It is often equivalent or interchangeably used with the term "fonds" in this dissertation.

Content: The information contained in archives.

Context: "The process of which the record is a part; the environment and web of relationships in which the document was created and used" (Williams, 2006, p. 8).

Context information: "Information describing the circumstances under which records … have been created and used" (Report from Toronto Archival Context Meeting, 2001, sec. Definitions and Uses, para. 2).

Element: A word, phrase, or group of characters representing a distinct unit of information that forms part of an area of formal description (Pearce-Moses, 2005, p. 113).

Encoded Archival Context (EAC): An XML (eXtensive Markup Language) standard used to markup archival context, including relevant information about the circumstance of record creation and use (Staatsbibliothek zu Berlin, 2017).

Encoded Archival Description (EAD): An XML standard used to markup archival finding aids including the hierarchical structure and descriptions of archival collections (Library of Congress, 2018).

Entity: A physical, digital, conceptual, or another kind of thing with some fixed aspects. It also "refers generally to all objects or agents of the system being analyzed" (Pearce-Moses, 2005, p. 148).

Event: "The (actual or contemplated) fact of anything happening; the occurrence of"; "Anything that happens, or is contemplated as happening; an incident, occurrence" (Oxford English Dictionary, 2014).

Fonds: The "whole of the records, regardless of form or medium, organically created and/or accumulated, used by a particular person, family, or corporate body in the course of that creator's activities or functions" (Bureau of Canadian Archivists, Planning Committee on Descriptive Standards, 2008, p. D-5).

Methontology: A systematic framework for developing new ontologies from scratch. It includes 1) the identification of the ontology development process; 2) a life cycle based on evolving prototypes; and 3) techniques to carry out management, development-oriented, and support activities (Fernández-López, Gómez-Pérez, & Juristo, 1997).

Ontology: An explicit formal specification of the conceptualization (Gruber, 1993).

OWL (Web Ontology Language): A family of knowledge representation languages for authoring ontologies. The languages are based on a description logic that supports ontology representation and reasoning (Hitzler, Krötzsch, Parsia, Patel-Schneider, & Rudolph, 2012)

Provenance: "the origin or source of something, or as the person, agency or office of origin that created, acquired, used and retained a body of records in the course of their work or life" (Millar, 2010, p. 98).

Resource Description Framework (RDF): A graph model for data interchange on the Web (Schreiber & Raimond, 2014).

Record group: "A collection of records that share the same provenance and are a convenient size for administration" (Pearce-Moses, 2005, p. 330)

Semantic Web: A Web of data that extends the World Wide Web through connecting, sharing, and reusing data across boundaries of software applications, websites, and people. It also refers to a set of technology standards that are used for describing, representing, and searching data on the data Web (W3C, 2009).

CHAPTER 2. LITERATURE REVIEW

This chapter reviews the literature related to archival descriptions and the Semantic Web to provide a foundation for this study. This chapter consists of three sections. The first section explains the definitions, features, and principles of archival descriptions. The major archival description standards and models are also explained. The second section reviews the definitions and features of the Semantic Web and ontology models. The third section explains the conceptual framework of the study.

2.1. Archival descriptions and archival principles

2.1.1. Definitions and features of archival descriptions

An archive is understood as a repository or collection of records with long-term value and defined as "document[s] regardless of form or medium created, received, maintained and used by an organization or an individual" (Cook, 1999, p. 5). An archival record is a record that exists uniquely in the world and contains "recorded evidence of an activity" (Shepherd & Yeo, 2003, p. 2). Archival records are primary sources that describe and record the direct or indirect interactions between the creators of records and the surrounding environment. This characteristic lends archives their indispensable value of being informational, evidential or both. Another view of records is that records are “persistent representations of activities or other occurrents, created by participants or observers of those occurrents or by their proxies; or sets of such representations representing particular occurrents” (Yeo, 2008, p. 136). The representational view of records focuses on capturing the dynamic features of records and emphasizing the activity (occurrent)-centred nature of records.

A record consists of content, context, and structure as its three fundamental components. Content refers to “the intellectual substance of a document, including text, data, symbols, numerals, images, and sound” (Pearce-Moses, 2005, p. 89). Context indicates “the process of which the record is a part; the environment and web of relationships in which the document was created and used” (Williams, 2006, p. 8). Structure indicates “the physical and intellectual characteristics that define how a document was created and maintained” (Williams, 2006, p. 8). These three aspects of a record are closely related to how a record is arranged and described. Records are often aggregated by hierarchical structures with descending levels of order (i.e., fonds or groups, series, sub-series, files or folders, and item). Each level reflects the creator of the fonds or group, functions or activities performed by the creator (series and sub-series), the record-keeping structure adopted by the creator or archivists (files and folders) and the actual materials kept (the item) (Millar, 2010).

To describe records in archives, the concepts and principles of archival descriptions have evolved since they were first introduced, according to how they have been used by archivists and society over time (Duranti, 1993). Archival description refers to “the process of establishing intellectual control over holdings through the preparation of finding aids” (Evans, Harrison, Thompson, & Rofes, 1974, p. 421). As information communication technology has an impact on how archival records have been described and organized by transforming from a paper-based practice to a computer-based approach, the definition of an archival description has also expanded into a broader view:

> “the process of capturing, collating, analyzing, and organizing any information that serves to identify, manage, locate and interpret the holdings of archival institutions and explain the context and records systems from which those holdings were selected” (Society of American Archivists Working Group, 1989, p. 422).

Miller (1990) further expanded this definition by emphasizing five types of information that archival descriptions can cover. He states that archival description is:

> "the process of capturing, collating, analyzing, controlling, exchanging, and providing access to information about (1) the origin, context, and provenance of different sets of records, (2) their filing structure, (3) their form and content, (4) their relationships with other records, and (5) the ways in which they can be found and used" (Miller, 1990, p. 7).

Among the five types of information, "(1) the origin, context, and provenance of different sets of records" refers to information about the creator, context, and provenance and could help facilitate the identification, management, and understanding of records.

Based on the examined definitions of archival descriptions, researchers seem to agree that an archival description is a process-oriented approach as well as "the product of such a process" (Pearce-Moses, 2005, p. 317). Similarly, in the General International Standard Archival Description (ISAD (G)), archival description is defined as a process of creation, and its outcomes are "the products of the process" (International Council on Archives, 2000, p. 10). Emphasis is given on being "an accurate representation of a unit of description" (International Council on Archives, 2000, p. 10).

The principles of archival description deal with content, context, and structure of archival records to gain physical and intellectual control over records, facilitating retrieval and understanding archives. The fundamental archival principles include "*respect des fonds*, provenance, and original order" (Evans, McKemmish, & Reed, 2009, p. 130). *Respect des fonds* refers to "archives from different creating agencies should not be intermingled" and "the original order in which materials were created and used should be respected" (Millar, 2010, p. 101). It is the combination of the principles of provenance and original order.

Provenance has two meanings: “the origin or source of something” and “information regarding the origins, custody, and ownership of an item or collection” (Pearce-Moses, 2005, p. 317). Provenance is defined as the “agency, institution, organization or individual that created, accumulated and maintained records … in the conduct of its business prior to its transfer to a record centre/archives” (Sweeney, 2008, p. 194) and often refers to the creators of the records. In general, the principle of provenance demands the preservation of the connection between the records and their creators, which includes three kinds of general information: individuals, families, and corporate bodies. Original order refers to "the organization and sequence of records established by the creator of the record" (Millar, 2010, p. 100). The principle of original order is to preserve the record keeping structure adopted by the creator of records. The principle of original order requires that records be maintained in their original order to reflect the creator’s functional and organizational contexts reliably.

Among the three fundamental components of an archival record (i.e., content, context, and structure), context is probably the most elusive concept. Context is considered a complicated concept because it is a “dynamic construct” (Greenberg, 2001, p. 257); thus, it cannot be separated from content. Dourish addressed that "contextuality is a relational property," and "an occasioned property," and it “arises from the activity” (Dourish, 2004, p. 5). Winograd (2001) emphasizes the relation between interaction and context by stating “because of the way it is used … features of the world become context through their use" (p. 405).

In the field of history, Berkhofer (1995) identified several types of context, including past context (actual context), documentary context (recorded in artifactual sources), and represented context (construction or interpretation of the inferred actual context). The identification of different contexts is important in understanding contexts since archival descriptions are representations of documentary context, reflecting actual context, and it is assumed that the roles of archivists are to bridge contexts by creating archival descriptions.

In addition, the importance of context is recognized by several researchers and archivists who support either the fonds-based or the series-based approaches (Davies, 2003; Hurley, 1998; Scott, 1966). Archival context can be presented differently in each approach. The fonds-based archival model was initially proposed in the "Dutch Manual" (Muller, Feith, & Fruin, 2003) and adapted in the Archival Group and the American Record Group. Under the fonds-based model, context was closely tied to one primary creator or organizational unit on which basic descriptive instruments were designed (Davies, 2003). Since contextual descriptions were traditionally considered as parts of archival descriptions, one contextual description was to be attached to the description of each fonds.

In Australia's series-based model, context controls four basic elements: organization, agency, family, and person (Scott, 1966). Hurley (1998) argues that the "series" system emphasizes describing both context and records entities, and the complex, dynamic relationships between them, which implies that the Australian series system may be more flexible than the fonds based approach to accommodate context and its representation.

Along with the concept of context, provenance is considered the immediate context of records creation and ambiance context as the broader context of records. The importance of provenance has been emphasized in archival science by several researchers (Duchein, 1983; Jenkinson, 1922; Millar, 2002; Nesmith, 2006; Schellenberg, 1956). Early on, Jenkinson (1922) stated that without providing provenance for records, archival records could not keep the authenticity of records, although the subjectivity of archivists may exist in the appraisal, arrangement, and description. Later, Schellenberg (1956) stresses that informational and evidential values are ensured by provenance. Duchein (1983) emphasizes that the contexts of provenance, such as the who, what, when, where, and why of record creation, as well as the qualities of the creator, need to be considered when determining a fonds. While Duchein's view of provenance is general, Millar (2002) argues that "[r]ather than limit provenance to creatorship, we should expand the concept to incorporate the spatial and temporal qualities of archeological provenance and artistic provenance" (p. 12).

Millar (2002) constructively expands provenance to include three separate components: the "creator history, record history, and custodial history" (p. 14). The creator history is "the story of who created, accumulated, and used the records over time," focusing on the creator of the records (Millar, 2002, p. 12). The functions and activities of the creator should be captured in this history. The records history refers to "the story of the physical management and movement of the records over time" (Millar, 2002, p. 13). This history should accurately record how records are created and used, whether they are lost, destroyed, enhanced, or altered. Millar also calls this the history of recordkeeping. The custodial history is "the explanation of the transfer of ownership or custody of the records from the creator or custodian to the archival institution and the subsequent care of those records" (Millar, 2002, p. 13). Nesmith (1999) proposes a broader view of provenance and stresses that "the provenance… consists of the social and technical processes of the records' inscription, transmission, contextualization, and interpretation, which account for its existence, characteristics, and continuing history" (p. 146). Furthermore, Nesmith (2002) suggests that an analysis of provenance should include an analysis of the "societal and intellectual contexts" in which a record was created. However, such "societal and intellectual contexts" may not be easy to identify with limited resources.

Archivists should pay attention to the articulation of archival requirements for capturing provenance information and be involved in the development of systems for managing electronic records. For example, Bearman (1995) advocates the importance of the functional requirements for capturing provenance information and integrating the requirements into systems. Cook (1997) also stresses that archivists should focus on the "functions and business processes of the creator that caused the record to be created" (p. 58).

In digital environments, the proliferation of electronic records has a direct and significant impact on provenance. Since archival principles have been developed with paper-based archival records, the provenance of electronic documents is different from the traditional way of managing provenance in paper records. Managing electronic records is considered one of the most important challenges for archives in digital environments (Hickerson, 2001). It is necessary to evaluate whether the principles of archives (i.e., provenance, original order, and *respect des fonds*) can be applied to digital archives in the same way. With the development of the Internet and Semantic Web, archival principles also need to keep up with these emerging digital environments in which users can contribute to the creation of electronic records and collaborate with other creators or users more easily than before. For example, in the digital environment, the physical order of records is not as important as that of paper-based archival records. The original order expands to "abstract conceptual order, order on the user interface of application software, and order in storage" (Niu, 2015, p. 62). With regards to provenance, the World Wide Web Consortium (W3C) defined it as "information about entities, activities, and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability or trustworthiness" (Groth & Moreau, 2013, para. 1). This definition has been expanded to focus on entities, activities, and creators when developing data as a unit of archive in the digital environments.

Based on the examination of archival principles and context, it is agreed that the context of archival records has been highly emphasized in different perspectives. Based on the archival principles, to operationalize archival principles in practice, archival description standards have been developed in archival communities. This will be explained in the next section.

2.1.2. Archival description standards

When developing archival descriptions, three types of archival description standards have been made, with a focus on data content, data structure and data value standards (Society of American Archivists, 1999). Each type of standard plays a different role in developing archival descriptions.

Data content standards "provide guidelines for creating descriptions and outline elements with their syntax and form" (Zeng & Qin, 2008, p. 317). The Describing Archives: a Content Standard (DACS) is a national standard in the United States that was officially approved by the Society of American Archivists (SAA) as a SAA standard in 2004 and revised in 2013. DACS provides an "overview of archival description (including the requirements for effective multilevel description) for archivists, outlines the elements that must be included at different levels of description, and describes how those elements should be implemented" (Kumar, 2011, p. 183). In addition, DACS integrates rules to describe the creators of archival materials, construct archival authority records, and record forms of names. DACS illustrates how rules and principles might be implemented in the MAchine Readable Catalogue (MARC) and data structure standards (Society of American Archivists, 2004). DACS has been appraised as a "forward-thinking and comprehensive response to changing information needs" (Whittaker, 2007, p. 98). It separates descriptive content from context and stresses the importance of a multilevel description. However, DACS still consists of many unstructured elements and the relations between content and context are not explicitly addressed.

The Rules for Archival Description (RAD) is another data content-based archival description standard in Canada, which is based on the core concept of "fonds." The first edition of RAD was released in 1990. A comprehensive revision was based on the Canadian-U.S. Task Force on Archival Description (CUSTARD) project draft and DACS. RAD provides two sets of rules for descriptions. The first set is for describing archival materials, and the other provides guidelines for creating headings and form names. Using "multi-level description", RAD determines what information an archivist should capture and present. Unlike DACS, RAD does not provide specific implementation suggestions and grants more interpretation power to archivists, which may cause inconsistency problems across activists and archives.

Data structure standards refer to “what data elements are included in a system” (Benson, 2009, p. 161) and they indicate “how compound objects are put together” (Zeng & Qin, 2008, p. 9). The Encoded Archival Description (EAD) is a data structure or data encoding standard. Technically, EAD is a metadata scheme based on eXtensible Markup Language (XML) for encoding archival descriptions. EAD is “a pioneer implementation of a generalized markup language into a metadata encoding standard for the library and archival community” (Zeng & Qin, 2008, p. 53). Although researchers agree that the structure that is captured in the EAD scheme does reflect the hierarchical structure in archival records (Pitti, 1999), over a decade later, some archives do not use it because it is too complicated to implement (Yaco, 2008).

The International Standard on Archival Description (General) (ISAD(G)) can be seen as another data structure standard developed by the International Committee of Archives (ICA). It intended to provide general guidance for describing archival materials for international archival communities. ISAD(G) provides a common denominator supplying a useful framework of principles and practice for many audience groups throughout the world and accommodates a range of needs in different practices, localizations, and customization with local standards. ISAD(G) provides structure and categorizes with six required elements (i.e., title, creator, dates, extent of the unit of description, level of description, and reference code) (International Council on Archives, 2000, p. 9), which results in different interpretations at the discretion of archivists and may cause ambiguity and inconsistency in archival practices in different countries and institutions.

Data value standards are “designed and used for controlling or restricting values in generating metadata statements, indexing, and retrieval” (Zeng & Qin, 2008, p. 320). The Library of Congress Subject Headings (LCSH) and Library of Congress Name Authority File (LCNAF) are examples of data value sets that help maintain consistency and standardize names and subjects. In archive communities, the creation of the International Standard Archival Authority Record for Corporate Bodies, Persons, and Families (ISAAR CPF, used as ISAAR (CPF)) was a major step towards the standardization of archival context information, in particular for creator descriptions, such as individuals, families, and corporate bodies. ICA published the first and second editions in 1996 and 2004, respectively. The second edition of ISAAR (CPF) consists of four sections of information, including identity, description, relationship, and control (International Council on Archives, 2004). In particular, ISAAR (CPF) covers a wide range of information about the creator, such as the date of existence, the functions and activities associated with the creator, the history of the creator, the internal structure (or genealogy) of the creator, and the overall context of the time in which the creator exists. However, it does not provide recommendations or information on implementation with regards to how the information should be formatted or stored with ISAD(G).

Based on ISAAR (CPF), the Encoded Archival Context-Corporate Bodies, Persons and Families (EAC-CPF) has been developed as an extension of EAD. EAC-CPF (referred to EAC hereafter) was released in March 2010. EAC addresses the description of individuals, families, and corporate bodies that create, preserve, use and are responsible for records in a variety of ways (Staatsbibliothek zu Berlin, 2017), and it complements and extends the EAD (Pitti, 2004). Like EAD, EAC provides a considerably well-defined structure and vocabulary. EAC is also an XML-based scheme with a nested hierarchical structure and formal specifications for encoding creator descriptions. The concept of relation has been introduced into EAC and used for capturing relations "between entities" (Pitti, 2004, p. 212). EAC's main advantage is to allow users to navigate "a web of relationships among records creators, functions, subjects and the materials" (Szary, 2006, p. 223).

However, since the EAC only focuses on authority information (i.e., person, corporate body, and family), it cannot express the more complex relationships among a record's creators, functions, subjects and materials, and a creator's functions and activities (Thurman, 2005). Furthermore, to assist in the design of EAC, a provisional document type of definition (DTD) called Encoded Archival Relations (EAR) was developed (Pitti, 2004), but only a handful of relations have been defined. In EAC, the creator description is addressed but other types of context information have not yet been captured. Without detailed context information, the linkages among related archival records are still limited.

As another venue, function is considered as "the whole of the activities that aimed to one purpose" (Duranti, 1998, p. 90). The importance of functions in the creation context of records is recognized and developed as the International Standard for the Description of Functions (ISDF). The concept of function, the models of functional descriptions and analysis of them are present in the archival appraisal stage, but they are rarely found in archival descriptions (Chaudron, 2008). Function is treated as an umbrella term and covers “not only function but also any of the subdivisions of a function, such as sub-function, business process, activity, task, transaction or other term in international, national or local usage” in ISDF (International Council on Archives, 2007, p. 7). Although the term function is used in ISAD(G), EAD and EAC, these standards did not provide guidance on describing and implementing functions. ISDF is a standard for describing a specific component “function” but limited to describe functions for corporate bodies. Since ISDF is a relative new standard, it has not been as widely as other standards (e.g., ISAD(G)).

Despite the advantages of the existing standards on archival descriptions, the current practices are not satisfactory to both users and professionals. From a user’s point of view, they often became “lost” and “overwhelmed” when they are faced with long container lists of many complex elements in archival description standards with a long list of files (Scheir, 2005; Yakel, 2004). Studies show that users dislike large blocks of text (Chapman, 2010; Scheir, 2005) because users experience “difficulty following large blocks of text in online finding aids” (Bahde, 2017, p. 485). Large blocks of text may result from paper-based archival descriptions.

In addition, an archival fonds/collection is often organized in a hierarchical structure with multiple levels (e.g., fonds, sub-fonds, series, sub-series, folder, file, and item). Users often have difficulty searching the hierarchical structures of archival collections that they are not familiar with (Chapman, 2010; Coats, 2004; Daniels & Yakel, 2010; Scheir, 2006). Furthermore, archival descriptions often provide contextual information at a higher level (e.g., fonds), and lower-level descriptions tend to contain limited information (Daines & Nimer, 2011). When users reach the lower levels of the hierarchical structure without knowing the related contextual information available at a higher level, this may bring only de-contextualized content to users (McCausland, 2011). Users also tend to be puzzled by unfamiliar archival jargon and labeling (Chapman, 2010; Daniels & Yakel, 2013; Lack, 2006; Prom, 2004). To solve this problem in archival descriptions, it has been suggested to use descriptive labels and terms that users can understand (Chapman, 2010; Nimer & Daines, 2008).

From a professional point of view, archivists struggle to deliver services when producing archival descriptions. For example, when using current standards, multiple provenances are not present in archival descriptions (Gilliland, 2012). With limited time and resources, archivists tend to produce archival descriptions in languages and manners that they are familiar with rather than users who use archives (Cox, 2007). It may be convenient for professionals to create archival descriptions with large blocks of text, without proper linkages and appropriate context. Moreover, although more and more archival information systems have become available to archivists when creating archival descriptions, supporting tools are still limited. It is very time-consuming and labour intensive to construct archival descriptions, link to related records, and update existing archival descriptions.

With the rapid development of Web technologies (e.g., Web 2.0, Linked Data, and the Semantic Web, etc.), in which web environments have become more interactive, and users' expectations have increased, the current practices of archival descriptions do not keep up with these changes (Light & Hyry, 2002; Trace & Dillon, 2012; Yeo, 2014). As search engines have become an indispensable tool for users to efficiently navigate the Web, users expect that the full text of archival descriptions and records should be searchable. Unfortunately, in reality, a large number of archival records and descriptions are still not searchable. In line with this situation, Daines and Nimer (2011) remarked, a "new medium requires a different approach and follows different practices" (Daines & Nimer, 2011, p. 9). Nielsen (1999) also recommended reducing text by at least 50% when moving from print to online environments and revising content for easier scanning. To improve the search capability of archival descriptions, they should be adapted to fit new technologies and meet users' expectations. In order to keep up with the Web environments, it is necessary to examine archival descriptions, especially archival context, and understand how to represent archival descriptions and context systematically. The next section explains various models to present archival descriptions and contexts.

2.1.3. Models related to archival descriptions and contexts

To capture, store, use and re-use archival descriptions, models for archival descriptions have been created and implemented in the field. Furner (2016) explains that the basic archival data model consists of two types of distinct entities: archival resource data and archival authority data. The Basic Archival Data Model is illustrated in Figure 2.1.

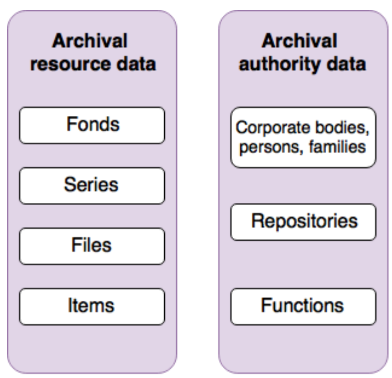


Figure 2.1 Basic archival data model (Furner, 2016, p. 45).

In Figure 2.1, archival resource data consists of fonds, series, files, and items, which correspond to the levels of the hierarchical structure of archival records. Archival authority data consists of corporate bodies, persons, families as creators, Repositories, and Functions as shown in Figure 2.1.This basic model is useful to identify types of data and understand the components of the archival model. However, it does not capture the relationships within and between each type.

There have been attempts to model archival descriptions in the relevant fields as well. In the field of records management, a high-level model has been proposed for defining and standardizing recordkeeping metadata (McKemmish, Acland, Ward, & Reed, 2006). This model provides a conceptual framework for presenting contexts and their interactive relationships and capturing the components of recordkeeping processes and their metadata, as displayed in Figure 2.2.

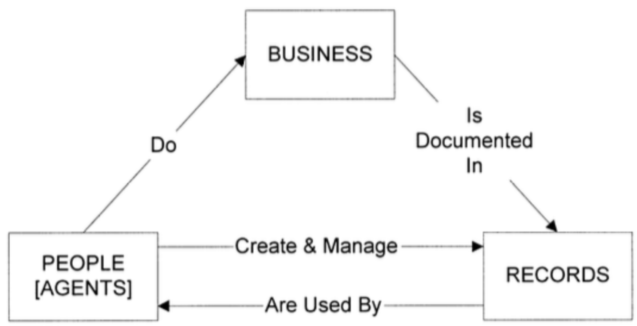


Figure 2.2 Recordkeeping Model (McKemmish et al., 2006, p.12)

The model consists of three main components: people [agents], records, and business. In business processes, people create and manage records in which the business processes are documented in those records. In other words, when people do business, people use the records that are created and managed by them. In comparison with the other two components (i.e., people and records), business is the main process component of this model. Rules, policies, and mandates regulate the process of business, record creation, and management. This model has been expanded to include a "mandate" entity for presenting the regulatory context that is established by people, governs the business process, and accounts for the execution of records.

Hurley (2009) adopted the conceptual model and further adjusted it into the deed, document, and doer model for archives, as displayed in Figure 2.2.

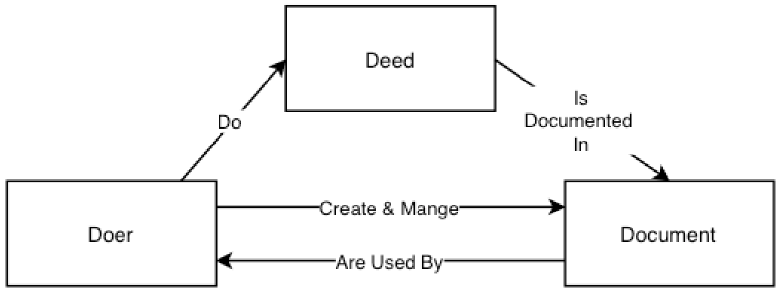


Figure 2.3 Deed, document, and doer model (Hurley 2009. p.1)

In Hurley's model, components are replaced: for business with deed, for records with document, and for people with doer. Document in Figure 2.3 refers to "the information or content that comprises the record" (Hurley, 2009, p.1). The content of deed must be "linked to some timebound event or circumstance in order to be a record." Specifically, deeds refer to "acts, actions, processes, activities, or functions" (Hurley, 2009, p.1); and doer is "the actor who undertakes the deed" (Hurley, 2009, p.1). The descriptions of deeds and doers are recorded in descriptions of documents. This model connects the three components (i.e., deed, document, and doer) in alignment with the recordkeeping model. The deed, document, and doer model is a high-level reference model for identifying the three components and capturing the relationships and interactions between them.

These three components of the deed, document, and doer model can be aligned with three international standards (i.e., ISDF is for deed, ISAD(G) is for document, and ISAAR (CPF) is for doer). More specifically, Hurley (2009) defined a deed as "a function, an activity or an action" and in his view a function is "a role or purpose, and an end rather than the means by which it is accomplished"; an activity is an application of "a Function enabling the outcome to be predicated in a specific instance"; and an action is a "step taken in pursuance of an Activity, a specific instance" (p.1). Hurley's model can act as an overarching model for covering the three standards. Activities are important building blocks of the model to link with function, action, and deed.

Although these three models identified the components (e.g., agent/doer, deed/business, document/record, and function) of archives, they did not capture other aspects (such as cultural, technological, and societal) of archival context. As another fundamental component of archives, context, or contextual information, has been defined (as explained in section 2.1.1, Definitions and features of archival descriptions) and been systematically presented through models. To understand the concept of context, the activity theory has attempted to model context (Huang & Gartner, 2009; Kaenampornpan & O'neill, 2004; Kofod-Petersen & Cassens, 2006). Activity theory explains that human activity is situated in historical and cultural contexts (Foot, 2014). The concept of activity refers to what subjects (e.g., people) do together (human activities) and are modified by both cultural and historical influences (Kaenampornpan & O'neill, 2004). The model of an activity system is constantly evolving through individual or collective actions and in response to systemic contradictions. Activity systems are multi-voiced and multi-layered to model activities undertaken by subjects with differing roles, positions, perspectives and objects (Yamagata-Lynch, 2010). According to Engeström (1987), an activity system consists of six core components: Subject, Object, Tool(s), Community, Division of labour, and Rules. These six components and relationships are illustrated in Figure 2.4.

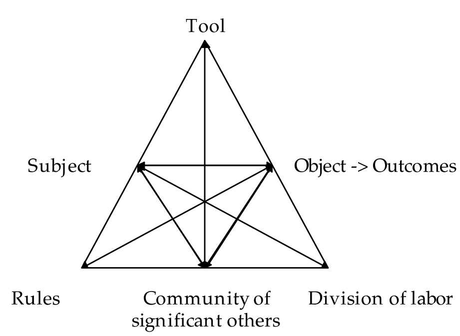


Figure 2.4 The model of activity systems (Engeström, 1987, p.78).

Among these components, the major ones include a subject (or actor), an object (a focal entity and the desired outcome), and the tool(s) that are employed by the subject to act on the focal object or pursue the desired outcome. Tools can be either material or conceptual, including language, protocols, scientific methods and models, and other forms of cultural artifacts such as hammers or computers. The community consists of the people who share an interest in the subject and have involvement with the same object. The interactions between the subject and the community that engages in a shared object can be thought of as the "communicative relations" of the activity (Engeström, 1999). The rules regulate the subject's actions toward an object and relations with other participants in the activity. The division of labour includes both the relatively horizontal division of tasks and the vertical division of power, positions, access to resources, and rewards (Engeström, 1987, 1990). Particularly, Rules, Community of significant others, and Division of labor are considered the social context (Huang & Gartner, 2009). Activity theory provides a basic framework for modeling social-cultural context in archives.

As contexts are indefinite (Duff & Harris, 2002) and limitless (Yeo, 2012), some researchers (MacNeil, 2005; Nesmith, 2006) stress that broader descriptive systems are necessary to enable the capture of the societal, cultural and historical context of archival records in archival descriptions. In line with this view, the concept of record has been further considered as "persistent representation of occurrents" (Yeo, 2008, p. 136). The concept of records is considered a social action (Foscarini, 2013) or information system (Lemieux, 2014). Several researchers advocate that archival context needs to be extended to include social and cultural aspects of context (Bastian & Alexander, 2009; Lemieux, 2014, 2015; Nesmith, 2006; Piggott, 2012; Yeo, 2007) so that archival context can be multi-faceted and complex (Foscarini, 2010; Lemieux, 2014; Niu, 2013).

Thus, to understand the cultural, social and technological aspects of context, activity theory can be adopted. Cole (1996) refers activity theory as Cultural-Historical Activity Theory (CHAT). An activity system is represented as a triangular model (Engeström, 1987) that consists of six components (i.e., Tool, Subject, Rules, Community of significant others, Division of labor, and Object). An activity system can capture a specific activity at a given time. The unit of analysis can be expanded from a single activity system to multiple, minimally two, interacting activity systems, which Engeström (2012) represented as third generation activity theory. Interacting activities indicate two or more activities with objectives that are ongoing. Kaenampornpan and O'Neill (2004) proposed modeling the history of context by using activity theory along with time, because each activity system is associated with a specific time as a series of activity systems. Activity systems analysis is a method, spawned from CHAT, for analyzing human interactions by identifying human activity (Yamagata-Lynch, 2010). Thus, activity systems analysis is adopted to model context in archives as interacting activity systems at a high level with the consideration of spactiotemporal features of human activities.

Several context models have been developed in the field of digital libraries and digital archives. When studying digital collections, Lee (2011) proposed a context model to capture the major contextual categories, which consists of nine classes: Object, Agent, Occurrence, Purpose, Time, Place, Form of expression, Concept/Abstraction, and Relationship. This model considers *Occurrence* as important as *Object* and *Agent*, and the importance of *Relationship* is also recognized. However, the nine classes are in a flat structure. This model does not provide a high-level view or details on how the nine contextual classes connect to each other.

Similarly, the Europeana data model (EDM) was developed for "structuring the data for ingesting, managing and publishing" contextual information (Europeana Project, 2013, p.4). EDM is built as a general model to "adopt an open, cross-domain Semantic Web-based framework" (Europeana Project, 2013, p.5) and features five classes entities (i.e., Agent, Event, Place, TimeSpan, and Concept) devoted to the representation of contextual information. It can accommodate object, contextual entities, and event-centered descriptive practices and be used as a denominator across cultural heritage communities.

Comisión de Normas Españolas de Descripción Archivística (CNEDA) (2012) in Spain proposed a conceptual model, including Records, Agent, Business, Mandate, Concept, Object or Event, and Place. This conceptual model was built on the Recordkeeping model (McKemmish et al., 2006). Relationships and attributes have been defined accordingly. In the model, functional context and regulatory context explicitly link to related Agent and Records so that the understanding of archival context expands from a functional context to include cultural and societal aspects of context. Thus, the Spanish model is a starting point that has built on archival theories. However, the Spanish model is limited because the contexts are narrowly defined. Furthermore, in this model, Concept, Object or Event refers to an "abstract notions or ideas, material things, actions or occurrences that are subject of records" as one entity (Comisión de Normas Españolas de Descripción Archivística, 2012, p. 15). These entities are so different from each other that they should not be amalgamated into one entity. In addition, time as an entity is not explicitly identified in this model.

The Finnish conceptual model for archival descriptions was developed in 2003 (National Archives of Finland, 2003). The model consists of 13 entities including Subject, Temporal Event, Lifecycle Event, Expression, Mandate, Past Lifecycle Event, Place, Planned Lifecycle Event, Function, Agent, Manifestation, Item, and Information Resources. Among the 13 entities, Agent, Function, and Information Resource are the core entities (National Archives of Finland, 2003).

Along with the 13 entities, 15 types of relationships are defined. This model is "based on the analysis of international and national standards" including ICA standards and ISO 23081:2 Metadata for Records, etc. Although the model intended to bridge "semantic consistency in description practices in cultural heritage sectors" (National Archives of Finland, 2003, p. 6), the details on the event-related entities (i.e., Temporal Event, Lifecycle Event, Past Lifecycle Event, and Planned Lifecycle Event) are not included and how events are related to Function and Mandate are not explained. No further development has been reported since the release of this model in 2003.

The International Council of Archives (ICA) aims to develop "a clear and comprehensive conceptual framework for archival descriptions" (Gueguen, da Fonseca, Pitti, & Grimoüard, 2013, p. 575). As a result, ICA presented the Records in Contexts (RiC) standard for archival descriptions in 2016, which consists of 14 entities (i.e., Record, Record Component, Record Set, Agent, Occupation, Position, Function, Function (Abstract), Activity, Mandate, Documentary Form, Date, Place, and Concept/Thing) (Experts Group on Archival Description, 2016). Properties of each entity have been presented with the 752 defined relationships, which identified a set of important elements in archival descriptions. Without a conceptual model explicitly defined and no detailed explanations of the selection of entities and relationships, the 752 relationships and entities do not present a functional and logical model that is built on archival principles; as Hurley (2017) criticized, "the draft is not … a conceptual model at all but a set of elements" (para. 3). In particular, Function (Abstract) and Function are two confusing classes in the RiC model since the differences between the two are not clear because Function (Abstract) seems unnecessary (Jones, 2017). Position and Occupation are more intuitive as properties instead of entities. Present and past tenses used in relationships are unnecessary (InterPARES Trust, 2016; Jones, 2017). More importantly, although there are many relationships defined, there is a lack of a core set of relationships to explicitly connect contextual information in archival descriptions. Thus, ICA's RiC model does not deliver a clear and comprehensive conceptual framework.

Based on an examination of the previous models, the existing standards and models have limitations in providing a conceptual model for guiding the creation and management of archival descriptions. In particular, there are not enough ways to explicitly connect contextual information in archival descriptions. This is resonant with what Niu (2014) points out by saying that "[c]urrent narrative descriptions … could be redesigned to be more structured for richer descriptions and improved searching due to their limitations" (p. 327). Although the Spanish model is a solid starting point, the RiC standard and Finnish model were developed with different views. The necessity for such an explicit and comprehensive conceptual framework for archival context should be addressed. The overarching comprehensive conceptual framework should be built on new understandings of archival principles, archival context, and practices.

In summary, this section has explained the definitions and principles of archives and archival descriptions. Archival descriptions are considered to be processes to establish administrative and intellectual control over archival materials and provide the structure, content, and context information in archives. Archival descriptions standards are explained, including data content standards (e.g., DACS and RAD), data structure standards (e.g., EAD and ISAD(G)), and data value standards (e.g., ISAAR (CPF)). To examine the models to present archival context, the Recordkeeping, Deed, Document, and Doer, Spanish, Finnish model, and RiC standard have been explained. The existing models have attempted to model archival context with entities, properties, and relationships, which are at the fonds/series level of current archival descriptions, but they are still not sufficient to explicitly present the contextual information for archival records. To better describe and present archival records, this study will explore the emerging semantic web and ontology technologies to represent archival descriptions, especially archival context.

2.2. Semantic web, ontology, and knowledge graph

To understand the formation of ontologies, this section explains the basic features of Semantic Web and ontology technologies. These have brought about new opportunities to present and model archival descriptions in a networked Web environment.

2.2.1. Definitions and architecture of the semantic web, ontology, and knowledge graph

Since the mid-1990s, the Web has provided a large amount of information and documents, most of which have been in HTML, PDF, and other formats. To add meaning to the information on the Web, the concept Semantic Web was proposed (Berners-Lee et al., 2001). The World Wide Web Consortium (W3C) defined the Semantic Web as "a common framework that allows data to be shared and reused across applications, enterprises, and community boundaries" (W3C, 2009, para. 1). It can "help machines understand more information on the Web and support richer discovery, data integration, navigation, and automation of tasks" (Koivunen & Miller, 2011, para. 3). In other words, the Semantic Web requires data to be explicitly presented, it makes relationships between data and provides a semantically rich description of information resources that can be linked, referenced and consumed by either people or software applications. Thus, the Semantic Web is set to transform a document based Web to a data Web (W3C, 2009).

Figure 2.5 shows the architecture of the Web platform for building the Semantic Web (Berners-Lee, 2000).

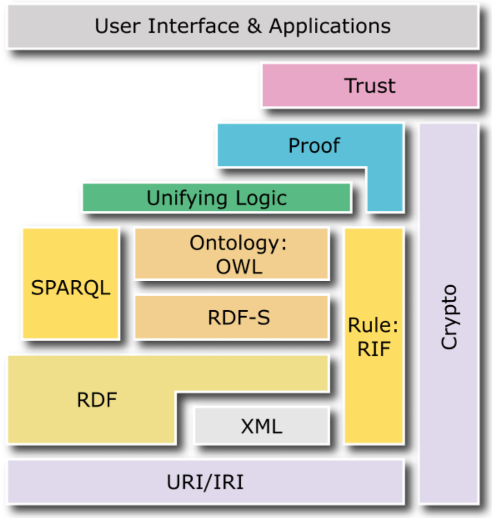


Figure 2.5 The semantic web stack (W3C, 2007).

In the making of the Web, the HyperText Transmission Protocol (HTTP), Transport Communication Protocol/Internet Protocol (TCP/IP), and Domain Name Systems (DNS) serve essential roles on the Web. Among the layers in Figure 2.5, the Resource Description Framework (RDF) is one of the most fundamental building blocks of making the Semantic Web. RDF is a general data model for data exchange and a language for representing information about resources on the Semantic Web (Manola & Miller, 2004). In RDF, Univeral Resource Identifiers (URIs) / Internationalized Resource Identifiers (IRIs) are utilized to describe resources with properties and property values in triple statements *(S, P, O)*, in which:

S is a URI and refers to the *subject* or *resource* of the statement.

P is a URI and is called the *property* or *predicate* of the statement.

O is either a URI or a *literal* and refers to the *object* of the statement.

An RDF statement can express a binary relation. For instance, (*sub, p, v*) is a statement indicating that *sub* is the subject or resource of the statement (e.g., https://en.wikipedia.org/wiki/Mark_Twain is a URI for Mark Twain), *p* is the property or predicate of the statement (e.g., http://purl.org/dc/elements/1.1/creator is a URI indicating him as a creator), and *v* may be literal or referring to an object (e.g., “Adventures of Huckleberry Finn” can be an object (another resource)). The statement “Mark Twain is the creator of Adventures of Huckleberry Finn” can be denoted in Figure 2.6 as follows:

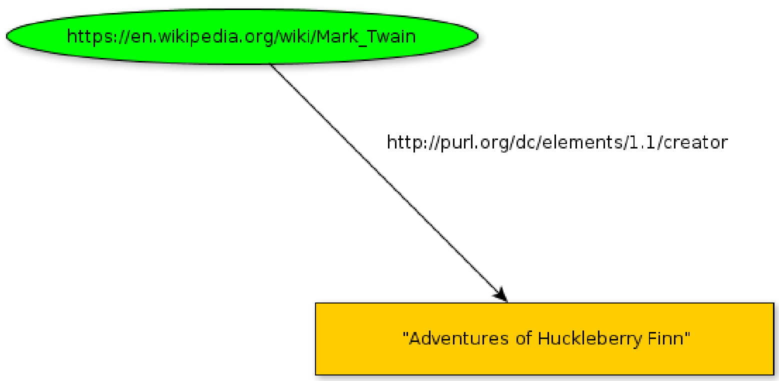


Figure 2.6 An RDF graph of the example.

Unlike other data models, RDF supports defining formal semantics. In predefined RDF semantics, entailments and corresponding RDF interpretations are connected (Hayes, 2004). For example, an RDF axiomatic triple *(rdf:subject, rdf:type, rdf:Property)* defines that *rdf:subject* is an instance of *rdf:Property* but not vice versa. The RDF data model requires that resources and properties be explicitly defined by using URIs/IRIs. This feature enables archival descriptions to be aggregations from other sources and builds well-defined internal structures for archival descriptions. This enables RDF to be used as the representation model for describing archival descriptions in this study. RDFS and OWL can further extend RDF to describe resources and properties (Matthews, 2005).

RDF-S is a schema language that purports “defining classes and properties that may be used to describe classes, properties and other resources” (Brickley & Guha, 2004, sec. Introduction; para. 6) and to “provide a simple reasoning framework for inferring types of resources" (Matthews, 2005, p. 5). RDF and RDF-S could be applied to transform EAD records into an RDF format. RDF and RDF-S emphasize semantic relations that are key to organizing and understanding archives. In the case of EAD, if an archival description of a record is stored as one document in EAD, its layout and structure remain limited within the record because EAD in an XML format addresses only the structure of EAD records without semantic relations. Furthermore, RDF and RDF-S require that the granularity of archival description should be at least at the RDF statement level, which ensures that archival descriptions have more granular data than documents.

As an example of adopting the RDF data model for describing and organizing records, a conceptual model of recordkeeping metadata is proposed (McKemmish et al., 2006) and later accepted as the International Organization for Standardization (ISO) 23081. The model consists of five entities: Agents, Records, Business, Mandate, and Relationship. The model adopts the RDF data model as a way to structure its metadata. However, in this model, RDF/RDF-S provides limited expressive power without more entailment rules, so it is unable to represent complex knowledge of archival description rules (Hitzler, Krötzsch, & Rudolph, 2009).

To understand ontology technologies, the concept of ontology has been used with a long history in relevant fields. In a general sense, ontology is situated in a non-philosophical context to address the problem of representing *machine understandable* information.

A widely accepted definition of ontology is given by Gruber (1993) who addresses that an "ontology is an explicit specification of a conceptualization" (p. 199). This definition indicates an ontology is an abstract and simplified view of the world (Gruber, 1993). More specifically, Uschold and Gruninger (1996) state that "an ontology is a shared understanding of some domain of interest" (p. 96).

An Ontology represents knowledge that can be computationally processed. The fundamental roles of knowledge representations are of a surrogate, "a fragmentary theory of intelligent reasoning" and "a medium for pragmatically efficient computation" (Davis, Shrobe, & Szolovits, 1993, p. 17). Ontologies can support richer and higher-level semantics because they can formulate a representation of a domain by identifying and defining concepts and the relations between them and support computational semantics. That is why ontologies are better than traditional ways of organizing concepts, properties, and relationships among concepts.

Ontologies can be grouped into lightweight (e.g., glossaries and dictionaries) and heavyweight ontologies (e.g., taxonomies, thesauri, and description logics) (Uschold & Gruninger, 2004). Heavyweight ontologies, often called formal ontologies, are usually described by knowledge representation languages such as OWL (Web Ontology Language). Heavyweight ontologies take advantage of the formal (i.e., rigid) definitions and reasoning power of knowledge representation languages to ensure a shared and common communication within one domain. In this research, the term "ontology" refers to an ontology in a general sense and formal ontology is used specifically for heavyweight ontologies. In addition, ontologies can also be grouped into three types: upper ontologies, domain-specific ontologies, and task/application ontologies (Guarino, 1998). These three types of ontologies serve different purposes: upper ontologies focus on high-level concepts across all knowledge domains (Phytila, 2002); domain-specific ontologies apply to specific domains only, and task/application ontologies apply to specific tasks or applications within domains. Among these three types, to better capture the processes of the creation and use of records, archival descriptions tend to be associated with specific domains, not general ones. Thus, ontologies in archival descriptions fall into the category of domain-specific ontologies.

OWL provides richer semantics than RDF and RDFS so that OWL classes, properties, the relations among them and constraints could be used to express complex knowledge. OWL is organized into three sub-languages: OWL Lite, OWL DL (Description Logic), and OWL Full, each with different expressive power (Patel-Schneider, Hayes, & Horrocks, 2004). These three sub-languages can be considered as extensions of RDF with different restrictions. For example, OWL Full is a language without expressive limitations but may not be computationally decidable, meaning that no reasoning software can perform complete reasoning for it. OWL DL has an expressive power and is decidable (i.e., complete reasoning is possible). OWL Lite offers a less expressive power but enough to model simple ontologies. The OWL 2 standard was released in 2009 to meet the information management needs of the real world (W3C, 2009). As OWL has the expressive power of RDF and the subtype semantics of RDFS, it can be used to model the logical representation of domain knowledge. By using OWL DL, the DL reasoning of ontologies can be utilized to detect inconsistencies and deduce implicit consequences from the explicitly represented knowledge (Baader, Calvanese, Mcguinness, Nardi, & Patel-Schneider, 2007).

Knowledge graph has gained popularity from the introduction of Google's Knowledge Graph (Singhal, 2012). Knowledge graph has a close association with Semantic Web technologies. A knowledge graph "acquires and integrates information into an ontology and applies a reasoner to derive new knowledge" (Ehrlinger & Wöß, 2016, p. 3). According to this definition, the Knowledge Graph is an application which includes collection or extraction pieces of information from different sources, integration of information from external sources into a knowledge-based system built on knowledge bases (e.g. ontology), and the generation of results (knowledge graphs) by a reasoning engine. Furthermore, knowledge graph applications leverage technologies developed for Linked Data and the Semantic Web. Thus, the Semantic Web could be "interpreted as the most comprehensive knowledge graph" (Ehrlinger & Wöß, 2016, p.3) and conversely "a knowledge graph that crawls the entire web could be interpreted as self-contained Semantic Web" (Ehrlinger & Wöß, 2016, p.4).

2.2.2. Ontology development processes

Ontology building is not a simple task as the complexity of ontologies is explained in the previous section. Since developing an ontology is a process that requires several activities, there are two types of ontology development processes. The first one is to build an ontology from a vacuum in the beginning and the second is to start with existing knowledge organization systems, such as controlled vocabularies, taxonomies, and thesauri or existing ontologies (Fernández-López & Gómez-Pérez, 2002). Editor Tools, such as Protégé (Stanford Center for Biomedical Informatics Research, 2016) and OBO-Edit (Day-Richter, Harris, Haendel, & Lewis, 2007), have been developed to facilitate the construction of ontologies. Tools for automatic ontology construction in natural languages, such as TextOntoEx (Dahab, Hassan, & Rafea, 2008) and GRAONTO (Hou, Ong, Nee, Zhang, & Liu, 2011), are available. However, semi-automatically and automatically generated ontologies still need domain experts to evaluate and verify them.

Broadly speaking, although an ontology development process may be arranged differently in different methods, it generally contains the following phases: (1) requirements and analysis; (2) design and implementation; (3) testing and validation; and (4) maintenance (Jones, Bench-capon, & Visser, 1998). Building an ontology involves many activities, such as environment scan, feasibility study, specification, conceptualization, formalization, implementation, maintenance, and use and each activity should be carried out very carefully (Fernández-López, Gómez-Pérez, & Juristo, 1997). In addition, knowledge acquisition, evaluation, and documentation should be performed during the entire life cycle of ontology development. Activities for building an ontology can be categorized into three types: (1) “ontology management activities” (e.g., scheduling, control, and quality assurance); (2) “ontology development-oriented activities” including pre-development (e.g., environment study, feasibility study), development (e.g., specification, conceptualization, formalization, and implementation), and post-development activities (e.g., maintenance and use); and 3) “ontology support activities” (e.g., knowledge acquisition, evaluation, integration, documentation, merging, alignment, and configuration) (Corcho, Fernández-López, & Gómez-Pérez, 2006, p. 10). The overall activities in the categorization vary depending on researchers.

There are several methods for building an ontology. To recognize the fundamental principles of ontology development, several approaches for building ontologies have been developed. Popular examples include methontology (Fernández-López et al., 1997), Model Driven Architecture (MDA) (Gašević, Djurić, & Devedžić, 2006), the Toronto Virtual Enterprise Method (TOVE) (Grüninger & Fox, 1995), and Enterprise Ontology (Uschold & King, 1995). Model Driven Architecture (MDA) is a software design approach for developing an object model as a ground by use of software systems (e.g., Unified Modeling Language) and then systematically transforming it to a formal ontology (Keshk & Chambless, 2008). The Toronto Virtual Enterprise Method (TOVE) takes defining motivating scenarios, informal competency questions, terminologies of the ontology, formal competency questions and the semantics and constraints by order (Grüninger & Fox, 1995). The contribution of TOVE is to use competency questions as a basis for defining the scope of an ontology that has been adopted in methontology (this is explained hereafter). The Enterprise Ontology method is a procedure used to identify terms and produce definitions. It also provides ways to handle ambiguous terms to reach a shared understanding of a common model. Each method has its advantages with a different focus.

Among these approaches to building an ontology, methontology is a systematic and comprehensive method for modeling a development process and builds ontologies from the beginning (Fernández-López et al., 1997). Methontology has been shown to be a comprehensive method implemented in many studies, for example, e-Government ontology (Dombeu & Huisman, 2011), Information Science ontology (Sawsaa & Lu, 2012), legal ontologies (Corcho, Fernández-López, Gómez-Pérez, & López-Cima, 2005), online learning (Guinebert, Yessad, Muratet, & Luengo, 2017) and chemical ontologies (López, Gómez-Pérez, Sierra, & Sierra, 1999). Cardoso's survey (2007) showed that methontology is the most widely used method (13.7%) to develop ontologies, while other methods are adopted comparatively less with On-To-Knowledge (7.4%) and Uschold and King's method (4.2%) (p. 87).

When adopting a methontology, ontology building is an evolving process that is comprised of multiple stages, including "specification, conceptualization, formalization, integration, implementation and maintenance" (Fernández-López et al., 1997, p. 35). A specification is producing a specification document written in natural language that may use a set of intermediate representations or competency questions. The specification stage should include the following information for conceptualizing the ontology: 1) the purpose of the ontology, 2) level of formality of the ontology, and 3) scope (Fernández-López et al., 1997). The second stage, conceptualization, is a process to structure the domain knowledge in a conceptual model based on specifications identified in the specification activity (Fernández-López et al., 1997). A complete list of terms including concepts, instances, verbs, and properties are identified to formalize the ontology. At the Formalization/Integration stage, existing ontologies are selected to fit the conceptualization. More importantly, the knowledge model is formalized. The next stage, implementation, expresses the ontology in a formal language. The formalization and implementation stages could be merged into one stage to ensure that there is no concept duplication, excessiveness, and inconsistent relationships. The final, maintenance, stage is when problems are fixed, and adjustments are made to the ontology.

As an example of adopting the methontology in the field, Sawsaa and Lu (2012) applied methontology to build an Information Science ontology confined to archival science, library science, and computer science at the specification stage. They divided the conceptualization into two parts: 1) build an information science glossary that includes terms with synonyms and acronyms and simple descriptions of terms; and 2) identify classes and relationships and construct a concept hierarchy out of the terms. At the formalization stage of building an information science ontology, the Protégé ontology editor was used to generate/formalize a knowledge model in Web Ontology Language (OWL). At the evaluation stage, the criteria (e.g., completeness, consistency, and clarity) was used to evaluate the information science ontology and fix inconsistent definitions or classifications and identify circular definitions.

Through this evolving process, concepts, entities, and relationships of an ontology become clear. Knowledge acquisition is accomplished via several methods, including content analysis, survey, interview or the focus group method. Taking each step in a series of stages, it is possible to explicitly express concepts, entities, and relationships among them and as a result, a domain ontology can be developed. In comparison to other existing methods, methontology not only recognizes the importance of knowledge acquisition (Gašević et al., 2006) but is also based on a real-world ontology building application (Fernández-López et al., 1997). Since methontology is suitable for building ontologies either from the beginning or by reusing existing ontologies (Gašević et al., 2006), employing methontology has the advantage of developing a representation of archival context. Therefore, this study adopts methontology to formalize the components of archival descriptions and to build an ontology for archival descriptions.

An event is an important concept in archives. The following section reviews event and event ontology building.

2.2.3. Event and event ontology

The concept of event has been examined since the time of Aristotle in several fields, including philosophy, history, linguistics, artificial intelligence, and others. In philosophy, Kim (1993) indicates that events are property instantiation; in other words, he considers history as smaller, interconnected event instances. Lewis (1986) sees events as a property of a spatiotemporal region. (?? Demerging) Galton (2008) addressed that since event, time, state, process and action are a cluster of concepts that are entangled with each other, it is difficult to define the concept of event from only one aspect. Furthermore, Galton (1995) groups time models into two kinds: 1) independent time models, where time exists independently of change, and 2) dependent time models, where time only exists, depending on the notion of change with which it is defined. Together with time, the resulting events are bound with one of these time models. In philosophy, events are grouped either by a three-dimensional view or a four-dimensional view. While the three-dimensional view refers to the three dimensions of Euclidean space with absolute time, the four-dimensional view sees time as a fourth dimension along with Euclid's three dimensions (Galton & Mizoguchi, 2009).

In the fields of linguistic and artificial intelligence (AI), some researchers realize the importance of precise analysis of the semantics of actions, events and the relationships between them (Allen, 1983; Kowalski & Sergot, 1986; McCarthy & Hayes, 1969). In AI, where "events as occurrences over intervals" (Bennett & Galton, 2004, p. 14) are well accepted, an event **e** can be associated with an interval **t** and a predicate **Occurs** denoted as **Occurs(e, t)**. Temporal relations between events can be described regarding relations between the intervals over which they occur in particular by the 13 interval relations: before/after, meets/met-by, overlaps/overlapped-by, starts/started-by, during/contains, finish/finished-by, and equals (Allen, 1983).

In addition, events can be structured. Kim (1993) considers events as objects in time with properties and events are structured by other objects. Objects often have clear "spatial boundaries and vague temporal boundaries; events, by contrast, would have relatively vague spatial boundaries and crisp temporal boundaries" (Casati & Varzi, 2015, para. 1). Events can link tangible objects or entities, such as people or recorded information, together and produce a web with an inter-connected picture that demonstrates what is involved in events and what their antecedents and consequences may be. More specifically, in archival records, there are several events related to the creation, use, arrangement, management, and description of records. In archival descriptions, events can be identified as activities of provenance, events of administrative history, and activities related to the creators of archives (Niu, 2014). As Yeo (2008) addressed, records are "persistent representations of activities or other occurrents" (p. 136) and records are social actions (Foscarini, 2013), to understand records, event-based information should be identified in archival descriptions and contexts of archival descriptions need to be captured to present records (Niu, 2014). This perspective reflects the activity (occurrent)-centred nature of records. In addition, object-centred approaches may not be a perfect fit for modelling event-based information. For the following description, "Madeleine Parent studied at McGill University from 1937 to 1940 and received her B.A. degree in 1940", object-centred approaches identify objects (e.g., Madeleine Parent, McGill University, B.A. degree). It is awkward to link those objects together if "studied" and "received" events are not included as part of the solutions. If "studied" and “received” events are modelled as properties, then each verb needs to correspond to a property that is difficult to manage. However, it is intuitive to connect objects through events. An education event here can easily link Madeleine Parent, McGill University, and the spatiotemporal information together.Thus, to present the activity of records, it is necessary to investigate what an event is and how an event can be formally represented in archival descriptions.

To model events, various ontologies have been developed in the past decade. The following list presents the nine most commonly used ontologies.

1) The International Committee for Museum Documentation (CIDOC) Conceptual Reference Model (CRM) (Doerr, 2003),
2) ABC ontology (Lagoze & Hunter, 2001),
3) Event Ontology (EO) (Raimond & Abdallah, 2007),
4) Linking Open Descriptions of Events (LODE) ontology (Shaw, 2010),
5) Simple Event Model (SEM) (van Hage, Malaisé, Segers, Hollink, & Schreiber, 2011),
6) F-Model ontology (Scherp, Franz, Saathoff, & Staab, 2009),
7) Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) (Masolo, Borgo, Gangemi, Guarino, & Oltramari, 2003),
8) Suggested Upper Merged Ontology (SUMO) (Pease & Niles, 2002), and
9) Basic Formal Ontology (BFO) (Grenon & Smith, 2004).

These ontologies can be divided into three groups: (1) domain-specific ontologies with event models, such as CIDOC CRM; (2) general event models, including ABC ontology, EO, LODE, SEM, and F-Model ontology; and (3) upper level ontologies with event models, such as DOLCE, SUMO, and BFO. While the five ontologies of the second group are relatively general and simple, the ontologies in the first and third groups are more complex as they have more classes, properties, and relations and an event model is a part of the entire ontology. The event-oriented models are characterized by common components, including time, place, object, agent, and event, as displayed in Table 2.1.

Ontology	Time	Place	Object	Agent	Event
CIDOC CRM	E52. TimeSpan, E2. Temporal Entity, E61. Time Primitive	E53. Place	E19. PhysicalObject	E39. Actor	E5. Event
ABC Ontology	Time	Place	Actuality, Abstraction	Agent	Event
EO	time:TemporalEntity	wgs84:SpatialThing	event:Factor, event:Product	foaf:Agent	eo:Event
LODE				foaf:Agent	lode:Event
SEM	sem:timeStampledAt	sem:Place	sem:Object	sem:Actor	sem:Event
F-Model	TimeInterval	Region	Object	Agent	Event
DOLCE	TemporalRegion	Region	Endurant, Continuant	Endurant, Continuant	Perdurant, Occurent
SUMO	TimeZone	GeographicArea	Object	Agent	Process
BFO	SpatialTemporalRegion	SpatialRegion	Continuant	Continuant	Occurent

Table 2.1 A list of event-oriented ontologies and mode.

As the first group, the CIDOC Conceptual Reference Model (CRM) is designed to describe museum artifacts and is used by libraries, archives, and museums (Gill, 2004). CIDOC CRM aims "to serve as the semantic glue needed to transform disparate, localized information sources into a coherent and valuable global resource" (Doerr & Crofts, 1999, p. 157). The CIDOC CRM has 86 classes and 137 unique properties in version 5.0.4 (CIDOC CRM Special Interest Group Working Group, 2011), and parts of them include event-related and class-based components. As shown in Table 2.1,two time components, Time-Span and Time-Primitive, are included. The core classes of CIDOC CRM cover Space and Time, Events, Material Things, and Immaterial Things. A subset of its classes and properties can be used for events. CIDOC CRM is a well-established model and created for serving as the basis of cultural heritage information. In addition, CIDOC CRM is a “property-centric” model that is built through a “property-driven” design process (Doerr, 2001). In CIDOC CRM, “Type” class is used for defining roles that may not be flexible because roles are often temporally bonded (e.g. Bill Clinton had the role of president from January 20, 1993 to January 20, 2001). Furthermore, CIDOC CRM does not directly support “function”, “mandate” types of abstraction, which are important for describing archival context.

In the second group, the ABC ontology was created for the Harmony Project, which aims to provide a conceptual basis for analyzing metadata vocabularies and instances and developing descriptive vocabularies and mappings between vocabularies (Lagoze & Hunter, 2002). The core intent of the ABC ontology is to model the creation, evolution, and transition of objects over time (Lagoze & Hunter, 2002). The ABC ontology consists of three primary categories under the class Entity (i.e., abc:Entity and abc is the prefix for the ABC ontology): abc:Abstraction, abc:Actuality, and abc:Temporality. The ABC Temporality category consists of three classes: abc:Action, abc:Event, and abc:State built on Situation Calculus (Lagoze & Hunter, 2002). The ABC ontology is simple and easy to understand. Because of its simplicity, many pieces need to be developed when applying it to archives.

The Event Ontology (EO), Linking Open Descriptions of Events (LODE), and Simple Event Model (SEM) are similar to each other. They are simple models without many restraints, which can facilitate interoperations. For example, the EO is built around six classes of which three classes (eo:Event, eo:Factor, and eo:Product) are defined in the EO (Raimond & Abdallah, 2007). The other three classes (foaf:Agent, geo:SpatialThing, and time:TemporalEntity) are reused directly from the Friend Of A Friend (FOAF) vocabulary (Brickley & Miller, 2014), the World Wide Web Consortium (W3C), Basic Geo vocabulary (Brickley, 2004), and the W3C Time

Ontology (Hobbs & Pan, 2006). The EO contains ten properties: agent/agent_in, factor/factor_of, literal_factor, place, product/produce_in, time, and sub_event.

The Linking Open Descriptions of Events (LODE) is a simple ontology (Shaw, 2010) which contains one class (lode:Event) and seven properties; two time related properties (lode:atTime and lode:circa); two place related properties (lode:inSpace and lode:atPlace); one property (lode:illustrate) to associate an event with a media object, and the other two properties (lode:involved and its sub-property lode:involvedAgent) used for relating instances of the Event class to instances of the Agent class. The LODE captures rudimentary properties and relationships related to events.

The Simple Event Model (SEM) is an event model with minimal restrictions and ontological commitment (van Hage et al., 2011). Its properties make it flexible for accommodating different views, and its classes are divided into three groups: core classes (sem:Event, sem:Actor, sem:Place, and sem:Time), types (sem:Typeclasses), and constraints (three kinds of sem:Constraints: sem:Role, sem:Temporary, and sem:View). Properties are divided into three groups: sem:event Property, sem:type, and a few other properties. The Role, one of 'Time-stamped entities', is treated as equally important as Actor, Event, Object, and Place. The Type class is also an interesting choice to define different types of Actor, Event, Object, Place, and Role, which can use a super-class/sub-class relation in ontologies. Roles and functions are mixed into one Role class. Spatial and spatiotemporal regions are hard to distinguish from each other.

The F-Model is a formal ontology built on a Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) (Masolo et al., 2003) that is based on the three-dimensional view (Scherp et al., 2009). By using a pattern-oriented design approach, F-Model can be used to represent arbitrary occurrences and model the different relations and interpretations of events

(Scherp et al., 2009). Since the F-Model is strictly defined as the OWL-DL compliant, the ontological commitment is higher than other ontologies (e.g.,EO, LODE, and SEM ontologies in the second group).

In the third group, the DOLCE, Suggested Upper Merged Ontology (SUMO), and BFO (Basic Formal Ontology) are fully developed as upper-level ontologies. The Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) is based on “a fundamental distinction between *enduring* and *perduring* entities” (Masolo et al., 2003, p. 10) and called “an ontology of *particulars”* (p. 8). In DOLCE Version 2.1, while endurants (enduring entities) are present at any time they are present, perdurants (perduring entities) extend in time by accumulating different temporal parts so that they are only partially present. For example, a person is an endurant but a person's life is a perdurant. The top-level categories are Endurant, Perdurant, Quality, and Abstract.

The Suggested Upper Merged Ontology (SUMO) is tightly built with the Standard Upper Ontology Knowledge Interchange Format (SUO-KIF) – a language designed for use in the authoring and interchange of knowledge (Niles & Pease, 2001). SUO-KIF terms can be individuals (e.g., Bill Clinton), or classes (e.g., Person). Relations and functions are defined as instances of the class of all relations and the class of all functions, respectively (Pease, 2011). SUMO consists of 25,000 terms and 80,000 axioms (Pease, 2018). In the concept hierarchy of SUMO version 1.8, two classes, Process and Object, are disjoint siblings of the parent class, Physical.

The Basic Formal Ontology (BFO) is intended to be the basic constituent of reality and can use both three and four dimension entities (Grenon & Smith, 2004). It focuses on the task of providing a genuine upper ontology, which can be applied to specific domains (Grenon & Smith, 2004). The BFO introduces a specific type of ontology of enduring entities (or continuants in time) called SPAN, and one of the occurrents called SNAP, which reconciles the three dimensional with the four-dimensional views. As "continuants are subject to constant changes and occurrents depend on continuant objects as their bearers" (Grenon & Smith, 2004, p. 72), occurrents change alongside continuants. While continuants refer to entities that have continuous existence and persist in time, occurrents refer to processes, events, activities, and changes (Grenon & Smith, 2004). Occurrents are four-dimensional and bound in time as "they occur in time and unfold themselves through a period of time" (Grenon & Smith, 2004, p. 72). In the BFO, the highest class is Entity, which consists of two classes: Continuant and Occurrent. While sub-classes of Continuant are Spatial Region, Role, and Function, the sub-classes of Occurrent are Temporal Region, Spatiotemporal Region, and Process. In BFO, the way dividing the Entity into Continuant and Occurrent is distinguished from those of SUMO (Physical and Abstract are two direct sub-classes of Entity) and DOLCE version 2.1 (Endurant, Perdurant, Quality, and Abstract are four direct sub-classes of Entity).

Considering the features of the listed ontologies presented in Table 2.1, they differ in their focus and approach. However, as these models treat events as objects at the same level, these models have advantages over those focusing on objects only.

A general principle of ontology development is to make use of existing ontologies whenever possible. To address the specific need of archival descriptions, most of all, a domain-specific event ontology needs to be developed and built on a well-established upper ontology. Archival descriptions have unique characteristics. For example, descriptions at the fonds level focus on certain events, such as the birth, death, or accomplishments of the creators of the fonds. The context information of archival descriptions is essential to understanding when, where, why and how archival records were created and used. A newly built event ontology should be tuned to capture their events enough to fully describe the characteristics of archival descriptions. Although all ontologies and models explained here can be used to model events, none of these ontologies is perfectly suited for archival descriptions. Modelling archival context requires not only conceptualization of events but other concepts (such as situation and function). The ontology or model to be chosen should be easily applicable and expandable. Among the listed models, CIDOC CRM and ABC Ontology were created in domains that are closely related to archives. However, CIDOC CRM and ABC Ontology do not align with other upper ontologies. CIDOC CRM is more complex than the ABC Ontology and there are more limitations to building upon it than ABC. The ABC ontology is simpler and more event-centric than CIDOC CRM. However, both CIDOC CRM and ABC Ontology are domain-specific ontologies, so they cannot support interoperability for ontologies at the same level as upper ontologies. In comparison with SUMO and DOLCE, BFO is widely used in the biological and medical communities and supports concepts such as "Role" and "Function" better than CIDOC CRM. Therefore, BFO is a suitable upper ontology to build a new event-based ontology for archival descriptions. This approach can provide benefits from the existing ontologies and be flexible enough to be applied to archival descriptions.

2.2.4. Applications of linked data to transform archival descriptions

To apply semantic web and ontology technologies into fields including archives, libraries, and museums, in recent years a new movement called *Linked Data* has emerged. Linked Data refers to "a set of best practices for publishing and interlinking structured data" (Bizer, Heath, & Berners-Lee, 2009, p. 1). Linked Data is a lightweight approach for data dissemination and mesh-up functions on Web environments to implement Semantic Web technologies in various fields.

Berners-Lee (2006) explained the four principles of Linked Data:

1. "Use URIs as names for things;
2. Use HTTP URIs so that people can look up those names;

3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL);
4. Include links to other URIs so they can discover more things" (para. 3)

URI (Universal Resource Identifier) is one of the most fundamental components of the semantic web and appears in each of the four principles. URIs support linking on the Web, which provides the foundation for solving linking archival descriptions within and beyond the Web.

Gracy (2015) identified several critical challenges for applying linked data to archival descriptions, including 1) differences in granularity in archival descriptions; 2) inadequacies of current encoding standards to support semantic tagging; 3) critical contextual information in unstructured fields; and 4) limited resources available for archives. Niu (2016) reviewed the current implementation status of the linked data development in the fields of libraries, archives, and museums and identified four types of linked data that are generated for archival materials: "1) archival descriptions; 2) archival authority files for corporate bodies, persons and families; 3) controlled vocabularies for subject indexing; and 4) content annotations" (p. 88).

As an example of using a linked data approach, the LOCAH (Linked Open Copac Archives Hub) project intended to transform EAD records to RDF formats, and to link converted RDF records with library bibliographic records and other related resources (LOCAH Project, 2011). Based on the EAD XML scheme, the LOCAH project team developed a data model for archival descriptions in EAD more precisely than existing practices. This project could break down EAD documents into more granular RDF statements and produce a transformation model for EAD.

As explained in the previous section, 2.2.2. Archival Description Standards, the Encoded Archival Context - Corporate Bodies, Persons and Families (EAC-CPF) has also been used to describe the structure and creator description in XML more specifically than before. In addition, Mazzini and Ricci (2011) converted EAC-CPF (Encoded Archival Context-Corporate, Person, and

Family) into two ontologies by using RDF (Resource Description Framework). Their approach is to convert EAC-CPF explicitly and directly into classes and properties two ontologies. The first ontology is a “different formalization of the XML schema of EAC-CPF standard” (Mazzini & Ricci, 2011, p. 74) and has three classes (i.e., element, attribute, and controlled_value) and fourteen properties (i.e., mayContainElement, containRequiredElement, hasAttribute, hasRequiredAttribute, mayContainValue, reference, isElementOf, isRequiredElementOf, isAttributeOf, isRequiredAttributeOf, isControlledValueOf, mayContainDatatype, diagram_ref, and occurrence). The second ontology contains ten classes (i.e., entity, person, corporateBody, family, controlArea, descriptionArea, nameArea, language, place, and relation) and twenty-six properties. There are five main clusters, including controlArea, descriptionArea, nameArea, entity, and relation. Among the five clusters, controlArea, descriptionArea, and nameArea are directly adopted as classes like containers or aggregations of information. Furthermore, information such as history, context, and bioghist are still in text/strings (known as literal) without a structured format. This approach directly translates EAC-CPF to ontologies that have not benefitted from ontologies, but bring problems (e.g., classes such controlArea, descriptionArea, and nameArea are questionable).

Park (2015) proposes a linked data approach to transform both ISAD(G) (General International Standard Archival Description) for archival records and ISAAR (International Standard Archival Authority Record) for authority records. Unlike the ontology of EAC-CPF (Mazzini & Ricci, 2011), Park (2015) converts the areas defined in ISAD(G) into object properties in RDF instead of classes. Park's model captures the hierarchical structure of archival descriptions, as illustrated in Figure 2.7.

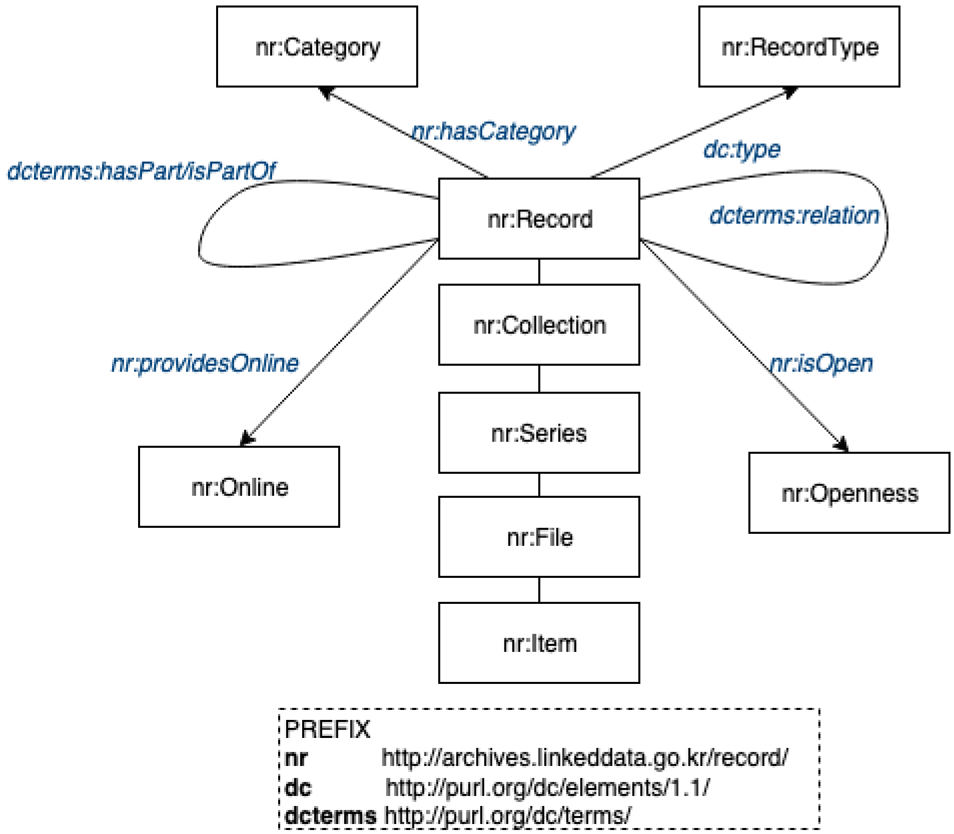


Figure 2.7 Model for archival data (this figure has been re-created) (Park, 2015).

In this model, the class *Record* is the center point. Collection, Series, File, and Item are defined as classes with Super/Subclass relationships (from general to specific). The *hasPart* / *isPartOf* relationships are taken from Dublin Core and used as object properties to express the hierarchies between records. A *relation* property (from Dublin Core) is to indicate the related records of the class Record. RecordType, Category, Online, and Openness are classes to support browsing. This model is a better way to structure objects in archival descriptions. However, this model is used only for transformation when migrating data into RDF. Thus, it is still limited to further model archival descriptions, especially archival context.

Another interesting project is Linked Jazz to leverage linked open data technologies to "represent and visualize the complex network of relationships held among jazz musicians as described in oral histories" (Thorsen & Pattuelli, 2016, p. 8). In particular, natural language processing (NLP) techniques were applied to analyze the transcripts of interviews of oral histories. Proper names of people are identified, and the basic relationships between people were identified (Thorsen & Pattuelli, 2016). This project is a good example to show the relationship between people and the content of archival records by using linked data technologies. The focus of the project was on musicians and relationships and could be extended to music pieces, performances, and recorded jazz music. However, no formal conceptual model was provided.

The "Out of the Trenches" is a proof-of-concept project by the Pen-Canadian Documentary Heritage Network (PCDHN). It aims to explore the application of linked open data to different forms of digital resources related to the First World War. Participating institutions include Canadian institutions, including the University of Alberta, McGill University, and others. A metadata model was developed with eight entities: Concept, Object, Event, Person, Family, Organization, Geography (or Place), and Asset (Information Resource). Existing authoritative thesauri (e.g., Library of Congress Subject Headings, Library of Congress Thesaurus of Graphic Materials, Bibliothèque Nationale de France (BnF)'s Rameau, UK Archival Thesaurus (UKAT), Canadian Subject Headings, and Government of Canada Core Subject Thesaurus) were utilized, which brought about advantages by leveraging the available knowledge organization systems in the library communities. This project shows, by using a sample dataset (i.e., about 1,200 RDF records/30,000 triples) (Pan-Canadian Documentary Heritage Network (PCDHN), 2018), how to link various resources (e.g., daily, newspaper, and government records) collected from various institutions (e.g., University Archives and National Archives) together. Unfortunately, the model that was developed in this project, especially the relationships among the proposed entities, was not explored further.

This section explained the features of the semantic web and ontology technologies, ontology development processes, and event ontologies. The previous or ongoing linked data projects for presenting archival descriptions provide ways to facilitate conversions on existing descriptions in EAD to RDF. Although the conversion of linked data has alleviated problems related to limited linkages, this approach does not deal with the full transformation of archival records encoded in the existing archival descriptions standards. The projects that are explained above have attempted to explore modeling archival descriptions in an EAD format or archival authority data into RDFs; they do not fully model archival descriptions in detail, and there are still limitations in capturing archival context and representing content and structure. Therefore, this research is designed to explore applying ontology technologies to archival descriptions and formally represent the archival context by using an event-based ontology model.

2.3. A conceptual framework for the research

A conceptual framework used for this study is based on theories from the following three disciplines: information science, computer science, and archival science. In the field of information science, as knowledge representation provides a basis for information retrieval (Weller, 2010), information scientists have studied methods of knowledge representation in taxonomy, classification, thesaurus, and ontology. In computer science, it is related to artificial intelligence that uses "formal symbols to represent a collection of propositions" (Brachman & Levesque, 2004, p. 4). Other relevant disciplines (e.g., linguistics, philosophy, cognitive science, educational psychology, etc.) also contribute to describing and defining knowledge and knowledge representations. The representation of archival descriptions is a combination of 1) knowledge representation of archival records; 2) the identification of entities from content, context, and structure of records; and 3) the modeling and processing of entities and their relations.

Several knowledge representation approaches have been developed in the field of information science, such as ontologies, thesaurus, taxonomy, and subject languages. Among them, ontologies provide more specific functionalities through formal conceptual descriptions of associations between entities and hierarchical structures. Ontologies can also use more formal symbolic systems than controlled vocabularies in a traditional sense.

Archival description is a process that describes archival collections in a specific way using knowledge embedded in textual descriptions. Archival descriptions are also special products that archivists develop to describe, organize, manage, preserve and provide access to archival records in collections (i.e., documented information). Context, content, and structure are understood to be critical to processing archival records in archives. As Hurley (2012) pointed out, existing archival descriptions are largely object-oriented in text-based forms, which makes it difficult to present processes, functions, and contextual elements of archival descriptions. Current archival descriptions are static representations of archival records since entities, relations, activities, and transactions are treated as objects in archival descriptions, there are limitations to describing existing archival descriptions. Thus, they are unable to explicitly capture the dynamic nature of archival records both in theory and practice. To overcome the limitations of archival descriptions, this study intends to explore and capture the dynamic aspects of archival context by identifying entities and relationships between entities described in archival descriptions. Adopting ontology technologies allows entities, events and activities to be identified and drawn from archival descriptions and to formalize them in an event-based ontology. Therefore, this study falls into the inter-disciplinary areas of knowledge representation, information science, and archival science, as illustrated in Figure 2.8.

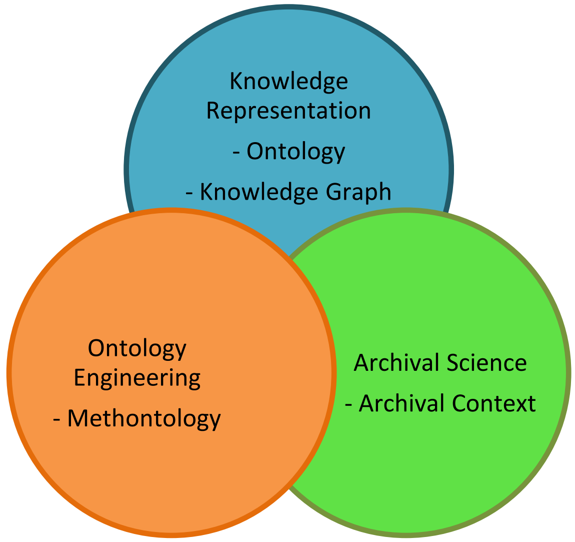


Figure 2.8 Intersections of the research in three areas.

Knowledge representation and archival descriptions are connected through ontology building by using ontology technologies. Among the several approaches to building an ontology, methontology is adopted in this study as a methodological framework that connects archival descriptions with knowledge representation. The methontology approach has been proven effective and widely used to construct ontologies in the relevant fields (Cordor, 2007). Methontology consists of specification, conceptualization, formalization, integration, implementation, and maintenance (Fernández et al., 1997) for building ontologies from the very beginning. Specification puts together a document that covers the primary purpose, the level of formality and the scope of the ontology. Conceptualization builds the domain knowledge in a conceptual model with specifications. In the formalization/integration stage, an informally perceived view of a domain is organized and converted into a semiformal specification (e.g., a glossary of terms in a domain, concept classification trees, and a binary relations diagram between concepts). Implementation is used to transform the conceptualization into an ontology language (i.e., Web Ontology Language). Maintenance validates the ontology, fixes problem areas of the ontology and keeps it updated with other ontologies associated with it. The following Figure 2.9 illustrates the stages of a methontology.

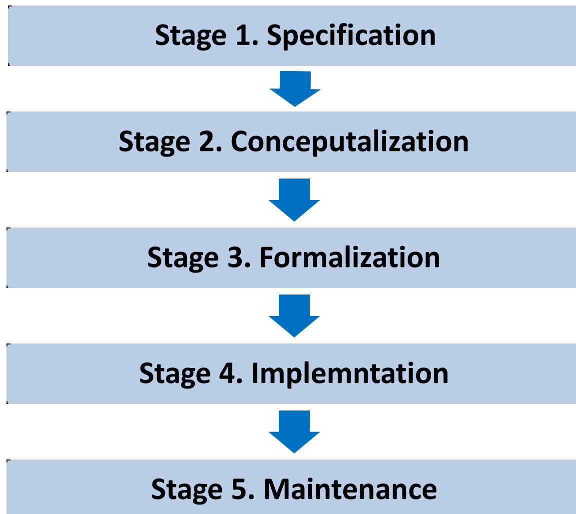


Figure 2.9 Ontology building stages of methontology.

The stages are not strictly linear, and some activities can last the entire process of building an ontology (e.g., documentation, integration). Evaluation may also occur at each stage. Knowledge acquisition activities, which take place throughout the specification and conceptualization/formalization stages, gather domain knowledge from archival descriptions, literature, and archival standards.

By taking advantage of methontology, this study will apply the ontology to archival descriptions and implement a new model of representing entities as events in archival descriptions. The outcomes of this study will add knowledge to the fields of knowledge representation and archival science by establishing a process-oriented ontology that recognizes the dynamic aspects of archival records.

In summary, this chapter has explained the major concepts of archives and archival descriptions, including context, provenance and archival principles. To understand ontology technologies, the basic features of semantic web, ontology, RDF, and ontology building processes have been explained. There are gaps in the existing knowledge concerning archival descriptions identified in this chapter. Although some projects have tried to utilize new methods (Bearman, 1989; Mazzini & Ricci, 2011; Millar, 2002; Park, 2015), there has not been a comprehensive model that fills the gaps. Thus, this study is designed to explore the representation of archival descriptions in the context of ontology technologies. The design and methods of the study are explained in the next chapter.

CHAPTER 3. METHODOLOGY

This study aims to investigate how archival descriptions can be represented in ontology technologies concerning archival context. This chapter explains the rationale of the research design, the choice of research methods, data sources and data analysis. In the first section, the rationale of research design and the selection of research methods are explained. The second section describes the data collection and analysis of the three studies. The third section discusses how the validity and reliability of the research are ensured. The fourth section summarizes the limitations of the research. The final section explains the ethical considerations when conducting the research.

3.1. Research design

To design a study, Crotty (1998) explains that research design elements indicate “aspects that inform a choice of approach” in research design (Creswell, 2003, p. 5). Based on the elements from the general to the specific, Creswell (2003) emphasizes that knowledge claims, strategies of inquiry, and methods are central to research design, which leads to research methods, such as quantitative, qualitative, and mixed methods. Then, the choice of methods can be translated into specific approaches (e.g., questions, theoretical lens, data collection, data analysis, write-up, and validation).

This research intends to examine archival descriptions, which can be created and represented as the objective facts with an emphasis on “accurate description”. Many studies in the field of archival studies, especially the ones related to archival descriptions, tend to take an objectivist stance. These archival descriptions reflect the informative and evidential values of archival records. They can be used to reflect the factual truth of archives.

As for research methods, three methods (e.g., qualitative, quantitative, and mixed methods) are commonly used in the fields of social sciences. In archival studies, qualitative methods are more often used than quantitative methods (Cox, 1994). A qualitative method is an "approach that centralizes and places primary value on complete understandings, and how people … understand, experience and operate" (Tewksbury, 2009, p. 39). Measurement for qualitative research occurs in the data collection processes, such as surveys, interviews, and historical studies. The data collected for qualitative methods often include written text, audio, video recordings, or visual images. Qualitative data analysis describes data in ways that "capture the setting or people who produced" (Schutt, 2012, p.321) on their own terms and interprets interrelated aspects of the setting, or people under investigation. In addition, qualitative data analysis is an iterative and reflexive process through the data collection process (Stake, 1995).

Quantitative methods use numeric measurements to develop specific variables or formulas as well as to test experiments and hypotheses (Creswell, 2003). Quantitative researchers design precise ways to measure objective facts. Quantitative researchers measure variables that are associated with research questions. Quantitative researchers develop techniques that can produce data in numerical values. Quantitative data analysis uses descriptive statistics to describe the distribution and relationship among variables and inferential statistics to "estimate the degree of confidence that can be placed in generalizations from a sample to the population from which the sample was selected" (Engel & Schutt, 2012, p. 371). Recognizing that each method has limitations, mixed methods can be used to supplement research findings through different instruments, facilitate development in the research process, and support the expansion of each method (Greene, Crarcelli, & Graham, 1989).

Regarding the multi-aspects of this research, this research adopts a multi-method for progressing all phases, complementing and maximizing the findings from each approach. More specifically, the two research objectives of this study are 1) to identify the elements for archival context and their relationships; and 2) to explore whether archival descriptions, especially regarding archival context, can be explicitly represented by using ontology technologies. In order to address the objectives, this study seeks to analyze the elements of existing archival descriptions regarding context and provenance and to represent archival context and re-build them in a model for archival context. Two research questions correspond to the two objectives, and the three methods were chosen for answering two research questions. The relations between objectives, research questions, and methods are shown in Table 3.1.

Objective	**Research Question**		**Method**
1. to identify the elements for archival context and their relationships;	RQ 1	RQ 1.1	Content Analysis
		RQ 1.2	
2. to explore whether archival descriptions,especially regarding archival context, can be explicitly represented by using ontology technologies.	RQ 2	RQ 2.1	Methontol ogy
		RQ 2.2	Focus Group

Table 3.1 Objectives, research questions, and methods

Research questions (RQ) 1 and 2 correspond to the two objectives. RQ 1 is formulated to address the first objective. In RQ 1, entities and their relationships of archival context in archival descriptions should be identified. To answer RQ 1, content analysis was selected. RQ 2 is further divided into two sub-questions. RQ 2.1 aims to explore the ways of representing archival context using ontology technologies. To answer question 2.1, a methontology was selected. RQ 2.2 is to evaluate whether the ontologies that are developed to address RQ 2.1 is perceived proper. A focus group method was chosen for addressing RQ 2.2. The following section explains the reasons and procedures for choosing the three methods.

3.1.1. Content analysis

To answer RQ 1, content analysis was chosen. Content analysis is a research method "for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use" (Krippendorff, 2004, p. 18). Content analysis is an unobtrusive method that can be used to analyze naturally occurring data. It assesses the use of words, phrases, or in-text relationships (Boettger & Palmer, 2010) and can be applied to draw trends, facts, and conclusions on the communicator, the text, and the sociocultural context. It has been widely adopted in social science, including library and information science and archival studies (Allen & Reser, 1990; Weare & Lin, 2000).

Content analysis can be quantitative or qualitative. Quantitative content analysis includes identifying meanings through valid measurement rules and making relational inferences with statistical methods (Riffe, Lacy, & Fico, 1998). In comparison, qualitative content analysis is for "the subjective interpretation of the content of text data through the systematic classification process of coding and identifying themes or patterns" (Hsieh & Shannon, 2005, p. 1278). Both methods of content analysis use research questions and examine language intensely for classifying large amounts of text that represent meanings of languages (Weber, 1990). Qualitative content analysis is guided by research questions through data gathering and analysis (White & Marsh, 2006). In qualitative content analysis, there are no existing hypotheses, but potential themes and other questions may arise from coding collected data (White & Marsh, 2006). In terms of data collection, qualitative content analysis collects naturally occurring data by using a purposive sampling technique for identifying complete, accurate answers to research questions. Data collection may continue throughout the study. In qualitative content analysis, coding is subjective, and a coding scheme is developed in the process of iterative reading to identify significant concepts and patterns.

Data analysis usually begins with the manual exploration of text and coding is the procedure used to find and label categories. Categorization is crucial for identifying specific language patterns of the text being examined. As a result of categorization, a coding scheme can be developed. A coding guideline is the collection of rules that the coders apply to the collected data (Saldaña, 2015). Quantitative content analysis uses a priori coding, where categories are set prior to the analysis to promote objectivity. Quantitative analysis uses descriptive statistics to test formulated hypotheses. Qualitative content analysis uses emergent coding where categories are established following a preliminary examination of the data, which can be subjective (Hsieh & Shannon, 2005). The processes of sampling and coding should be carefully taken to ensure the validity of a content analysis study. To strengthen the content analysis method, it can be combined with other research methods, such as interviews and focus groups (Kohlbacher, 2006).

Content analysis has its limitations. For example, biases may be introduced through the selection and sampling of data. As the development of the coding scheme and coding always involves interpretation, there are risks of overlooking what is not said in a particular text or omitting some significant text (Rose, Spinks, & Canhoto, 2015).

In this study, RQ1 aims to examine existing archival descriptions in order to identify the entities and their relationships among the identified entities. Since RQ 1 aims to find "what" entities are used (not "why") in archival descriptions, qualitative content analysis is suitable for this research. Qualitative content analysis aims to reveal and contextualize existing archival descriptions (the texts) produced by archivists (one of communicators) to the users (the other communicator). In the qualitative content analysis, archival descriptions have gone through three important steps (i.e., sampling, coding and analyzing). In the sampling step, the samples of archival descriptions in this study were retrieved at the fonds level. In the coding step, archival descriptions were coded manually by the researcher. In the analyzing step, the researcher analyzed the coded data and presented the results.

3.1.2. Methontology

To answer the first question of RQ2, methontology was chosen. As explained in detail in Chapter 2. Literature Review, Section 2.2.2 Ontology development processes, methontology is a systematic, well-structured, and sophisticated method for ontology construction, which provides a set of guidelines and techniques throughout the whole ontology development process (Fernández-López et al., 1997). It has been considered one of the most adopted methodologies in ontology developments by several researchers (Corcho et al., 2005; Park, Sung, & Moon, 2008). Methontology consists of detailed and well-defined activities that make it easier to carry out when building an ontology. Methontology has roots in main activities that are defined by the Institute of Electrical and Electronics Engineers (IEEE) software development process (IEEE, 1995). Methontology is composed of management, development, and support activities (Corcho et al., 2005). Management activities are for planning and quality assurance (Corcho et al., 2005). The development activities are specification, conceptualization, formalization, implementation, and maintenance (Corcho et al., 2005). The support activities consist of knowledge acquisition, integration, evaluation, documentation, and configuration management (Corcho et al., 2005). In addition, they suggest an iterative and incremental approach to developing an ontology. For example, knowledge acquisition – one of the support activities – is carried out throughout the development process and associated with other activities in groups.

Methontology also covers extensive, detailed tasks of each activity in an ontology development process. It supports the simultaneous use of top-down and bottom-up techniques, which helps developers effectively reduce mistakes (Park et al., 2008). More importantly, methontology is domain independent, so it enables the construction of an ontology at the knowledge or conceptual level. Thus, methontology has been regarded as one of the mature ontology development methodologies by several studies (Brusa, Caliusco, & Chiotti, 2008; Corcho, et al. 2003; Cristani & Cuel, 2005; Park et al., 2008, Al-Baltah, et al., 2014; Yu & Luo, 2012). In addition, among ontology construction methodologies, methontology has the highest adoption rate (13.7%) (Cardoso, 2007). Based on these evaluations of methontology, this study adopts methontology as a primary approach in building an ontology in the field of archival descriptions. Through knowledge acquisition techniques, methontology enables data (e.g., glossaries, relation tables, and concept dictionary) to be generated for the next phase of conceptualization. With acquired knowledge, tasks can be carried out in each phase and move to the next phase in the entire processes until the ontology is complete. Thus, methontology is a logical, systematic, and accountable method for building an ontology of archival descriptions.

3.1.3. Focus group study

User studies have been considered important to the field of archival science. In the 1980s, Conway (1986) first addressed a comprehensive framework of user studies in archives with several cases on how to study users in archival institutional settings and proposed a direction for archival user studies. To conduct user studies, research methods include surveys, interviews, and focus group studies (Rhee, 2015). In Web environments, where users and their behaviors change, user studies in archival science should change accordingly, with an emphasis on users' interactions with archival information systems.

Rhee (2015) grouped research topics of archival user studies into three general categories: “information needs, information seeking, and information use” (p. 32). The first category of information needs studies concerns the subjects including the archival users' research trends, interests, and the information sources they use (Rhee, 2015). The second category, information seeking studies, focuses on the information seeking behaviour of users (e.g., search strategies, using access tools, interaction with archivists) (Rhee, 2015). The third category of information use studies deals with the use of archival materials (e.g., citation patterns) (Rhee, 2015). In this study, the second sub-question of RQ 2 aims to reveal archival users’ perceptions of ontology representation in archival context. This question relates to the second category of information seeking and use because the representation of archival context in ontology forms influences and changes information seeking and use. For example, Duff and Stoyanova (1998) conducted a focus group study to obtain users’ opinions on the content and format of displays in archival information systems. The findings of the study demonstrate that users have problems interpreting information regarding physical description, dates of creation, and professional jargon such as “fonds,” while they consider information such as scope and content, abstract, and call number as essential for users (Duff & Stoyanova, 1998). As RQ2.2 aims to elicit responses from different user groups on the representation of archival context by using ontology technologies, a focus group method was suitable for answering RQ 2.2.

Focus group is a research method that “collects data through group interaction on a topic determined by the researcher” (Morgan, 2006, p.121) and allows participants to "react to and build upon the responses of other group members" (Morgan, 1997, p. 16). The focus group method typically has five distinct features, including 1) a group of people, 2) the participants possess with certain characteristics, 3) generating qualitative data, 4) having a focused discussion(s), and 5) having a topic of interest (Krueger & Casey, 2009). A focus group can provide "direct evidence about similarities and differences in the participants' opinions and experiences" (Morgan, 1997, p. 8). As focus groups encourage interaction among both participants and the researcher, who is the moderator, the individual who conducts a focus group needs to be careful to avoid bias from a dominant or opinionated member (Krueger & Casey, 2009). However, since a focus group typically is composed of a small number of people, this may limit the generalization to a larger population (Morgan, 2006).

In this study, a focus group method is adopted because it was necessary to identify users' responses and interactions to review a new proposed model in a research setting and recognize any flaws or problems with the model. Rhee (2015) grouped the different types of users in archival user studies in which two major participants of focus groups are commonly used: users with basic knowledge of archives (e.g. students, researchers, and genealogists) and professionals (e.g. archivists, librarians, experts). Based on the design of Rhee (2015), this study adopted two kinds of focus groups, a user group and a professional group.

In terms of the number of participants in each focus group, Krueger and Casey (2009) suggest, "the ideal size of a focus group for most non-commercial topics is five to eight participants" (p. 67). Based on the qualitative nature of the focus group and the study by Kruger and Casey (2009), the researcher aimed to recruit at least five to six participants for each group and ended up with six participants in each group. One participant in the professional group voluntarily withdrew from the study after the focus group session so that eleven participants were kept in total.

The research settings for organizing and conducting the two focus groups are designed to be consistent in order to ensure the reliability and validity of the study.

3.2. Data collection and analysis

As per the selection of research methods to answer the research questions, which have been explained in Section 3.1 Research Design, the detailed data collection and data analysis methods are summarized in Table 3.2.

RQ	Studies	Data sources	Data analysis
RQ 1.1 and 1.2	Study 1	Archival descriptions collected from archival institution sites	Qualitative content analysis for open coding and categorization
RQ 2. 1	Study 2	The results of Study 1 and archives related glossaries, thesauri	Identification of entities and their relations through methontology
RQ 2. 2	Study 3	Two focus group studies	Qualitative content analysis for coding and categorization

Table 3.2 A summary of data sources and analysis.

This research consists of three studies. As shown in Table 3.2, Study 1 uses qualitative content analysis to answer RQ 1. Study 2 uses methontology to develop an ontology in RQ 2.1. Study 3 uses a focus group study to address RQ 2.2. As the three studies used different data sources, data analysis methods are therefore different and depend on the raw data collected from each study.

3.2.1. Study 1

Study 1 sought to answer RQ 1 and focused on examining existing archival descriptions. A two-step procedure was developed to collect archival descriptions. This includes 1) the selection of two lists of archival institution sites, and 2) the retrieval of archival descriptions from the sites chosen in step 1. This procedure was tested before the data collection. Step 1 carefully selects data sources, while step 2 collects data consistently. Collected archival descriptions were cleaned, coded, and analyzed.

3.2.1.1. **Study 1 - data sources and collection**

The protocol of Study 1 includes 1) identification of data sources and 2) data collection. The details of the protocol are explained in the following paragraphs.

First, archival institution sites were identified from: 1) the implementor listing site of the Society of American Archivists (SAA) (Society of American Archivists, 2010) and 2) the repositories of primary sources site (Abraham, 2015). The SAA is one of the major institutions that has developed the Encoded Archival Descriptions (EAD) standard and is currently responsible for maintaining it. It is one of the most authoritative sites that contains the EAD implementor listing page. The page is a valid source for locating archival institutions that have implemented EAD, with more than 80 university archive sites.

As the secondary source, in order to complement sample sites from the SAA implementor listing site and to avoid sampling biases, if any, the repositories of the primary sources site was sought to collect additional archival descriptions. The repositories of the primary sources site list were chosen because they contain a comprehensive primary sources list available and have been used in other archival studies (Zhang & Mauney, 2013). The site contains "over 5000 websites describing the holdings of manuscripts, archives, rare books, historical photographs, and other primary sources for the research scholar" (Abraham, 2015, para. 1) and is organized by geographical region. Since archival descriptions are written in English in North America, which is the scope of this study, the selection of this site was an appropriate addition to complement the archival institution sites of the SAA list. The names of the sites and their URLs were listed by regional categories in alphabetical order. In this study, three regional categories have been chosen, including 1) the Western United States and Canada, 2) the Eastern United States, and 3) Canada. Then, the archival institutions listed in the three categories were retrieved.

The names of archival institution sites and their URLs were retrieved from the two lists, and two spreadsheets were created with the names of the archival institutions, the Uniform Resource Locator (URL) of the institution site, and a unique number assigned to each institution site. The original order of the institution sites in the list is intact, but each institution site has a unique row number in the spreadsheet. While the SAA EAD implementor spreadsheet has 89 institution sites numbered from 1 to 89, the repository of primary sources spreadsheet contains 2,837 institution sites, numbered from 1 to 2837. Then R, which is a "programming and language environment for statistical computing and graphics" (The R Foundation, n.d., para. 1) was used to generate two sets of random numbers. In order to select 40 institutions from the two lists, each set had 20 numbers that correspond to the numbers in the spreadsheets. Since each number could be considered as representing each institution site, 40 institution sites (20 institution sites from the SAA and 20 institution sites from the repository of primary sources) were selected. Each selected institution site was checked manually to make sure: 1) the URL of the site is valid and 2) the site contains archival descriptions. If the URL of the site was broken or archival descriptions were not available on the site, another institution's site was chosen randomly from the corresponding list and then the same procedure was followed until 20 institutions' sites were selected from each list. The quality of the repository on the primary sources list was not as good as that of the SAA EAD list. In the first round, only 10 out of 20 institutions' sites on the primary sources list were valid and archival descriptions could be found. To supplement 10 institutions' sites, in the second round, 20 institution sites were taken from the list again using the same selection criteria and procedures as described in the previous step. Then, only 13 institutions' sites were identified as valid. Among the 13 sites, 10 sites were randomly selected and added to the final list in Set 1. Set 2 contains 20 institutions' sites from the SAA EAD list. In this way, the list of 40 selected archival institution sites was established.

The second step was to retrieve archival descriptions from the 40 selected archival institutions' sites. For each institution site, three archival descriptions were randomly retrieved and saved in their original format mostly in HTML. Among these 40 institutions, since the technical infrastructure and the structure and design of each archival institution site vary, only 27 archival institution sites contain a complete list of archival fonds (e.g., Acadia University Archives, Rocky Mountain Online Archives, etc.). The remaining 13 archival institution sites did not display a full list of archival descriptions (e.g., Smithsonian Institution Archives, Duke University Archives). Instead, they provided a search interface, where the researcher could search archival descriptions with search terms. In the study, in order to bring up archival descriptions at the collection/fonds level, search terms such as "archives", "collection", "paper", "record", "fonds", "online archives", and "online collection" were used to search and thereby receive the search results. The number of search results was listed in a spreadsheet, ranging from a single result to a large number of results. For example, 55,000 results were returned by searching "collection" at the Smithsonian Institution Archives. Among the search results, three archival descriptions were randomly taken from 13 institution sites, producing 39 archival descriptions. Thus, there ended up being a total of 120 sample archival descriptions.

Most of the archival descriptions retrieved were in HTML format. Some archival institution sites provided archival descriptions in multiple file formats (e.g., HTML, PDF, and XML) and archival descriptions stored in different formats were retrieved. At this step, archival descriptions were stored in their original format without any conversion or transformation. Archival descriptions retrieved in HTML format were converted into plain text format. If some archival descriptions are in multiple formats, the format with the most comprehensive contents was taken and then converted to plain text format. During the conversion process, the information that is not relevant to archival descriptions was discarded, such as page numbers in PDFs, various headings H1 (i.e., Heading 1), H2 (i.e., Heading 2), and H3 (i.e., Heading 3) in HTML, and formatting information in PDFs.

3.2.1.2. **Study 1 - data preparation and analysis**

After the data collection process, 120 plain text files in EAD were ready for data analysis (e.g., coding) using qualitative content analysis. The retrieved data has been through a primarily manual process (i.e. data preparation, coding, and analysis), which was conducted by the researcher from May 24, 2014 to November 7, 2014.

Because this study does not have a pre-defined coding frame, a coding scheme had to be developed inductively for the collected archival descriptions. In general, inductive category development is a process in which the researcher creates the criteria for the category analysis, codes through the text, and formulates new categories based on the criteria.

Since a de facto standard TimeML is available, which is a Markup Language for Temporal and Event Expressions for annotating, extracting and reasoning about events in computational linguistics, the coding frame (i.e., events related parts) of this study was developed based on a coding guideline for TimeML (Saurí et al., 2006). According to TimeML, the unit of analysis is smaller than a sentence. One sentence may consist of several spans of text as entities (i.e., events) that need to be annotated.

During the process, entities were examined, and the categories were revised continuously. The reliability of the process was also checked in a feedback loop (Mayring, 2000). This method is called the constant comparative method which "compares each text assigned to a category with each of the texts already assigned to that category" and integrates categories and their properties through the process (Zhang & Wildemuth, 2009, p. 312). This study adopts the constant comparative method for coding in this study. Manual content analysis interprets categories in the text.

In order to ensure the validity and reliability of the research, a four-step protocol was developed for identifying the entities of archival descriptions and examining their specific relationships in terms of archival contexts. The four-step protocol is listed as follows:

1) Identifying general entities in archival descriptions
2) Identifying sentences in archival descriptions
3) Identifying granular entities in archival descriptions
4) Categorizing events in archival descriptions

Each step in the procedure is explained in the following sections.

Step 1: Identifying general entities in archival descriptions

Archival descriptions, especially archival contexts, contain information related to the creator(s) of archival materials, relevant people, organizations, and societal background. The entity here can be considered a container that comprises many sentences. Each sentence consists of smaller entities, including events. Entities are grouped by the original elements (or tags) that are assigned by whoever created the archival descriptions. In this sense, from general to specific, archival descriptions are categorized into entities, sentences, and events. The percentage of distribution can reveal the general composition of entities in archival descriptions. Thus, archival descriptions, especially archival contexts, consist of several entities, such as abstract, administrative history, biographical information, custody history, other notes, related information, and scope & content. Entities are also called elements in EAD (Encoded Archival Description) version 2002 and EAD 3.

Regarding the two sets of sample archival descriptions that are collected for this research, each description has been further dissected into general entities by abstract, administrative history, biographical information, custodial history, notes, and scope & content.

Step 2: Identifying sentences in archival descriptions

A sentence is a linguistic element and refers to "one or more main clauses, corresponding to units in which in written language are bounded by the punctuation mark" (Preisler, 1997, p. 24). To identify sentences, punctuation marks, especially the period ("."), are used. Although a period is used in some cases (e.g., Dr., Jr., etc.), computer algorithms can achieve high accuracy with an over 95% precision ratio and 90% recall ratio to detect sentences in natural language application programs (Agarwal, Ford & Shneider, 2005) or they can be manually identified. In this research, sentences are automatically identified and then manually reviewed.

Step 3: Identifying Granular Entities in Archival Descriptions

RQ 1.1 identifies entities in archival descriptions, especially in archival contexts. In general, an entity is defined as "a physical, digital, conceptual, or another kind of thing with some fixed aspects" (Growth & Moreau, 2013, p. 24). In a narrow sense, entities refer to "all objects or agents of the system being analyzed" (Pearce-Moses, 2005, p. 148). Relations refer to linkages between entities and can be treated as a particular type of entity. General entities (e.g., scope & content, abstract, administrative history, custody history, related information, biographical information, and notes) of archival descriptions are widely used in describing archival collections in archives, as explained in the previous section. General entities consist of granular entities and relations.

The manual process used to identify granular entities (including events) is explained in the following example. Details are given in APPENDIX I. Archival descriptions coding guidelines.

> [Example] The vast majority of the collection was donated to the College of Staten Island by James L. G. Fitz Patrick's stepson, Harold J. Smith, in 2000. A small number of news clippings were donated by Peter Spiridon, a former colleague at Staten Island Community College. James L. G. Fitz Patrick also donated several of his research publications to the College Archives prior to his retirement in 1976.
> (Source: Set 2, Institution #6, Archival Description #2).

a) Identifying and locating events described by verbs and nominalized verbs

The example above contains the three sentences that are identified by carrying out step 2 of the content analysis procedure and listed below:

Sentence 1: The vast majority of *the collection* **was donated to** *the College of Staten Island* by James L. G. Fitz Patrick's stepson, *Harold J. Smith*, in *2000*.

Sentence 2: A small number of *news clippings* **were donated by** *Peter Spiridon*, a former colleague at Staten Island Community College.

Sentence 3: *James L. G. Fitz Patrick* also **donated** several of *his research publications* to *the College Archives* prior to his **retirement** in 1976.

The identified verbs/verb phrases are in **bold,** including "was donated to" in Sentence 1, "were donated by" in Sentence 2, and "donated" in Sentence 3. A nominalized verb "retirement" is identified in Sentence 3.

b) Identifying other phrases associated with each verb

Since English is an SVO (Subject-Verb-Object) form, each verb in an SVO structure is associated with a subject and object (Tomlin, 1986). Subject and object are treated as granular entities in this study. Phrases directly associated with each verb are identified. In Sentence 1, there are four phrases directly associated with the verb in *italic*, including "the collection", "the College of Staten Island", "Harold J. Smith", and

"2000". In Sentence 2, "news clippings" and "Peter Spiridon" are directly related to the verb. In Sentence 3, "James L. G. Fitz Patrick", "his research publications", and "the College Archives" are directly associated with the verb ("donated"). There are two additional entities "James L. G. Fitz Patrick" (i.e., referring to "his" in "his retirement") and "1976" related to the nominalized verb ("retirement").

c) Marking appropriate spans of text to represent the phrases

For instance, "the vast majority" in Sentence 1 and "a small number" in Sentence 2 were marked as a feature of the collection. They are not directly associated with the verb in the sentence. In Sentence 1, "the collection" is considered as an appropriate span of text. In Sentence 2, "news clippings" is an appropriate span of text. In Sentence 3, "research publications" is an appropriate span of text. Regarding time, "in 2000" and "in 1976" are appropriate spans of text because they provide information about when the verb action occurred. Uncertain information is sometimes used in archival descriptions. For example, in Sentence 3, the donation happened sometime before 1976 (i.e., "prior to his retirement in 1976") and the phrase is an appropriate span of text for the verb "donated". However, "in 1976" is the appropriate span of text for the nominalized verb "retirement".

d) Assigning appropriate entities to the phrases

Events and relations are considered as special entities and identified among other granular entities in each sentence. In Sentence 1, the text "donated" was identified and indicated that a donation event happened ("something occurred"). A donation event was coded between two entities: "Harold J. Smith" (a person) and "the College of Staten Island" (a cooperate body). The event is also associated with other entities:

time (2000), artifact (collection). Meanwhile, a kinship relation (stepson) was identified between two entities: “James L. G. Fitz Patrick” (a person) and “Harold J. Smith” (a person). Similarly, events can be identified along with other entities in Sentence 2 and 3. In Sentence 2, the donation event was initiated by “Peter Spiridon” (a person) who is a formal colleague (relation) at “Staten Island Community College” (a corporate body). Their relation is further identified as a social relation. The entity “news clippings” falls into the general entity “Artifact”, which is not specific enough. A new type of entity called “information artifact” was created to further categorize the entity. In Sentence 3, “Research publications” is categorized as “information artifact” or “scholarly publication” as a more specific sub-entity under “information artifact”. Although “the vast majority” and “a small number” provide some information about the collection, they are not marked as a separate entity at this stage because they are considered as features of entities.

Three sentences can be dissected into granular entities, as illustrated in Table 3.3. The text of the right column in Table 3.3 contains phrases directly excerpted from the sentences. The column entity is divided into two columns: (general) entity and granular entity (if applicable), respectively. For the sake of simplicity, the features (e.g., “the vast majority”, “a small number”, and “several”) of entities are not presented in Table 3.3.

<table>
<tr><td rowspan="9">Sentence 1</td><td colspan="3">The vast majority of the collection was donated to the College of Staten Island by James L. G. Fitz Patrick's stepson, Harold J. Smith, in 2000.</td></tr>
<tr><td colspan="2">Entity</td><td rowspan="2">Text</td></tr>
<tr><td>General</td><td>Granular</td></tr>
<tr><td>Event</td><td>Donation event</td><td>donated</td></tr>
<tr><td rowspan="2">Agent</td><td>Cooperate Body</td><td>the College of Staten Island</td></tr>
<tr><td>Person</td><td>James L. G. Fitz Patrick, Harold J. Smith</td></tr>
<tr><td>Time</td><td></td><td>2000</td></tr>
<tr><td>Artifact</td><td></td><td>Collection</td></tr>
<tr><td>Relation</td><td>Kinship</td><td>Stepson</td></tr>
</table>

	Agent	Person	*James L. G. Fitz Patrick, Harold J. Smith*
Sentence 2	A small number of news clippings were donated by Peter Spiridon, a former colleague at Staten Island Community College.		
	Entity		**Text**
	General	***Granular***	
	Event	Donation event	*donated*
	Agent	Person	*Peter Spiridon*
	Artifact	Information artifact	*news clippings*
	Relation	Social Relation	*A formal colleague*
	Agent	Person	*Peter Spiridon*
		Corporate Body	*Staten Island Community College*
Sentence 3	James L. G. Fitz Patrick also donated several of his research publications to the College Archives prior to his retirement in 1976.		
	Entity		**Text**
	General	***Granular***	
	Event	Donation event	*donated*
	Agent	Cooperate Body	*the College Archives*
		Person	*James L. G. Fitz Patrick*
	Time	Interval	*prior to his retirement in 1976*
	Artifact	Scholarly publication	*research publications*
	Event	Retirement event	*Retirement*
	Agent	Person	*James L. G. Fitz Patrick*
	Time		*1976*

Table 3.3 Identification of granular entities with the sample sentences

Step 4: Categorizing events in archival descriptions

The procedure to identify events is similar to that of identifying granular entities and assigning them to event types. To explain the procedure for identifying events, the previous procedure for identifying granular entities can be extended with one more step to assign appropriate event types as follows:

1) Identifying and locating events described by verbs and nominalized verbs
2) Identifying other phrases associated with each verb
3) Marking appropriate spans of text to represent the phrases
4) Assigning appropriate entities to the phrases

5) Assigning appropriate event types to the events and linking events with related entities.

In the following example, four events have been identified, including three donation events as verbs and one retirement event as a noun.

Sentence 1: The vast majority of the collection *was donated to* the College of Staten Island by James L. G. Fitz Patrick's stepson, Harold J. Smith, in 2000.

Sentence 2: A small number of news clippings *were donated by* Peter Spiridon, a former colleague at Staten Island Community College.

Sentence 3: James L. G. Fitz Patrick also *donated* several of his research publications to the College Archives prior to his *retirement* in 1976.

From Steps 1 to 4 above, events were identified in each sentence by types of entities. In Sentences 1 and 2, two donation events can be identified. In Sentence 3, a donation and retirement event can be identified. A category of events can be established. Every event is checked against the category. If an event can fall under the category of events, the event is marked with the proper category. If not, a new category is created for the event. After the completion of coding, a categorization of events is generated.

3.2.2. Study 2

3.2.2.1. **Study 2 – data sources and collection**

Study 2 further examines the archival context in order to identify entities and their relationships from archival context information. These are used in building an ontology model. Entities and relationships are expressed in a glossary of terms. This study identifies, defines, and describes the entities and their relationships for use in building a model. As an ontology building method, methonology is used to build an event ontology for archival context.

The results of the qualitative content analysis are the primary data source in Study 2. In addition, existing thesauri and glossaries are supplementary sources for developing ontologies, including SAA Archival thesaurus (Pearce-Moses, 2005), InterPARES thesaurus (The InterPARES project team, 2001) and Australia Functions Thesaurus (National Archives of Australia, 2016). These three thesauri are used as they are the most well-known, popular and reliable in the field of archives in English speaking countries. Study 2 uses methontology, which is composed of multiple stages, including specification, conceptualization, formalization, integration, implementation, and maintenance for building ontologies as mentioned in the previous section (Fernández et al., 1997, p. 35). Through the specification, conceptualization, and formalization stages, terms from the data sources were compiled into a glossary of terms, and then a data dictionary provides concepts and relationships to the ontology.

3.2.2.2. **Study 2 - data analysis**

In this study, among the stages of the methontology, the first three stages (e.g., specification, conceptualization, and formalization) consist of core tasks for developing the ontology. The identified categories, entities, and relations are used to draw specifications and concepts in the first stage. The terminology, objectives, granularity level and scope of the ontology are identified at the specification stage as well. At the conceptualization and formalization stages, a complete glossary of terms was developed and then the terms, categories, and relations are turned into concepts, attributes, relations, and instances.

At the integration stage, some elements of upper ontologies (e.g., DOLCE and BFO) and other related ontologies have been explored. Based on an examination of existing event-related models (e.g., SEM, EVENT, LODE, CIDOC CRM) as explained in Chapter 2, section 2.2.3 Event and event ontology, the BFO ontology (*Basic Formal Ontology*, 2011) was adopted and reused with the conceptualization. The adopted ontology also incorporated other existing ontologies, such as FOAF (Friend of a Friend) (Brickley & Miller, 2014), Relationship vocabulary (Davis, 2010) and Dublin Core terms (Dublin Core Metadata Initiative, 2012). The detailed procedures of building an ontology are explained in CHAPTER 4. THE FINDINGS OF STUDIES 1 & 2. At the implementation stage, the ontology and vocabularies from the conceptualization stage are coded in RDF and OWL formats.

3.2.3. Study 3

3.2.3.1. **Study 3 - data sources and collection**

Study 3 aims to evaluate a proposed event ontology for archival context and identify problems and issues in the proposed event ontology. This study takes advantage of a focus group method in order to elicit feedback from users and professionals about the existing archival descriptions and the event-based ontology model developed in Study 2.

The focus group method was adopted in Study 3. In order to effectively ask users for their understandings of the proposed event ontology model, in Study 3, two groups of participants were recruited, including 1) a group of users (e.g. novice archivists; and 2) a group of professional archivists. The focus group sessions were held in two locations in Canada. The research settings where the two focus groups were organized and conducted were identical in order to ensure the reliability and validity of the study.

The participants of the first group were users and graduate students who have chosen the area of archives as their stream in graduate studies in information studies and would like to be professional archivists after graduation. These participants had an interest or background in understanding archival materials, archival descriptions and metadata. An invitation letter (see APPENDIX II. Letter of invitation) was sent to the e-mail list of the Master of Library and

Information Studies (MLIS) students who were graduate students at the School of Information Studies, McGill University in Montreal, Quebec. Based on their responses, six volunteer students were invited to join the focus group study on a first come, first served basis. For the first group, the focus group study was held on December 8, 2014, in Room 312 in the School of Information Studies building, McGill University. At the beginning of this focus group session, a moderator, who was the researcher, described the research objectives and provided a consent form (see APPENDIX III. Sample consent form) to each participant. The moderator answered participants' questions regarding the consent form and the study. After participants voluntarily signed the consent form, a questionnaire (see APPENDIX IV. Sample questionnaire) was used to collect some background information about the participants. Then, the archival description sheets (see APPENDIX V. Three sample archival descriptions for focus group discussions) were distributed to participants. The sheets contained three archival descriptions at the fonds level that were randomly selected from the 120 sample archival descriptions collected in Study 1. These descriptions are displayed both in traditional archival descriptions and an event-based ontology model. The sample descriptions included different types of creators in archival descriptions, such as individuals, families, and cooperate bodies. The researcher explained both the descriptions and proposed models to those who did not understand the technical details of the descriptions and the ontology model. All participants reviewed the three archival descriptions of the existing archival descriptions and the three event-based ontology representations, respectively. The moderator asked participants a list of questions (see APPENDIX VI. Focus group questions) that focus on comparing the proposed model and existing practices in archival descriptions. The moderator encouraged participants to express and discuss their thoughts, opinions, and observations about the advantages and disadvantages of both the current archival descriptions and the proposed model.

A semi-structured interview with participants followed. The session lasted about one hour and 15 minutes and was recorded using a digital audio recorder. The data collected from the first focus group was securely stored for further analysis.

For the second group of professional archivists, the invitations were sent to the Canadian archival listserv (arcan-l) to solicit their participation. Based on their responses, archival professionals who work in libraries, archives, and museums were selected. Participants of this group work in archives. Their responsibilities involve creating, organizing and managing archival descriptions in archives. Six participants participated in the study, but one participant withdrew from the study after the session. The focus group study of the second group was held on July 15th, 2015 at room LI5012 in the Chancellor Paterson Library, Lakehead University in Thunder Bay, Ontario. The focus group session was held in an identical manner with the first group. The moderator was the researcher, who asked participants a series of questions (see APPENDIX VI. Focus group questions). The session lasted approximately 60 minutes and was recorded using a digital audio recorder. The data collected from the second focus group was securely stored for further analysis. Each participant of both sessions received cash compensation for their time and contribution.

In both sessions of the focus group, participation was voluntary. Any identifying information on participants was coded and remains anonymous. All data from this study is stored on a secure server at McGill University for seven years and can be accessed by the researcher and advisor only. The findings of this study are available to participants upon request.

3.2.3.2. **Study 3 - data analysis**

To analyze the results of the two focus group studies, two types of data (the questionnaire and audio recordings) were collected. Data from the questionnaires were stored in a spreadsheet, and descriptive statistics were used to reveal the group composition of the focus group study. Audio recordings were played back using Audacity software and transcribed in text format and saved into textual files. The transcriptions include the number of participants, time, and actual content. After the researcher finished the transcriptions, the transcriptions were rechecked by the researcher for clarification and verification. Since time is included in the transcriptions, each sentence can be pinpointed using Audacity and played back. The second transcriber, who is a Master's student in public health from Lakehead University with experience transcribing audio recordings, checked the two audio recordings against the transcribed text to make sure that the text was transcribed exactly as in the audio recordings.

Thematic coding is utilized to identify the themes of the data. Since codebook development is one of the most critical steps in a thematic analysis, the transcribed texts were re-read and analyzed by segmenting and coding them into categories, types, and relationships. Coding analysis is a process of making sections of text, giving them labels or names, and also identifying when to use and not to use code (Holloway & Wheeler, 2010). A coding procedure for this study was developed as follows to 1) read through transcripts and then make some notes afterward; 2) read through them again and list possible codes while reading; 3) compile a list of codes; 4) review codes by deleting duplicated codes, merging codes, and adding code; and 5) link between codes. Applying codes to transcripts is a process of asking a consistent set of questions and refining the codebook.

The procedure to identify emergent themes in the analysis of focus group studies were developed including the following four steps:

1) Initial coding
2) Identifying emerging themes
3) Identifying superordinate themes
4) Summarizing themes

Step 1. Initial coding

The transcripts of the focus groups' comments and discussions were read carefully and meticulously for coding. According to Smith et al. (2009), during an initial coding stage, comments can be categorized as descriptive (e.g., the content or subject of talk), conceptual (e.g., more interpretative comments), and linguistic (e.g., language use). The descriptive and conceptual comments are the focus of this study since the linguistic features of the comments are outside of the scope of this study. While the descriptive comments tend to review the transcript mostly at face value, the conceptual comments involved more reflection or abstraction from the researcher. Thus, the interpretation took place at two levels. On the first level, the participants' points of view became clear, and then, on the second level, themes emerged. The participants' understanding or conceptualization of issues on archival descriptions could also begin to be considered. Throughout this process, the initial coding had to be closely matched to the participants' original accounts. As the analysis progressed, common and important themes emerged.

Step 2. Identifying emerging themes

This step is a process to link individual "parts" to the "whole" within the hermeneutic circle with a focus on the discrete sections of transcripts. When analyzing transcripts for their narratives, it is important to address newly emerging themes driven by the comments of participants and relate the emerging themes to the research questions in this study.

Step 3. Identifying superordinate themes

This step attempts to examine the relationship or connections of the identified themes with each other to identify superordinate themes. A superordinate theme is a construct that usually appears in most participants' accounts but can be represented in different ways. Table 3.4 displays a sample analysis to demonstrate the inclusion of emerging themes within superordinate themes.

Superordinate themes	Themes	Transcript	Initial notes
Relationships and their representation	The relationship between tags, event, person Representing in ontology Relationships among people Representing a certain type of relationship	(From Group 1) It depends on how you initiate such tasks because nowadays in *semantic web* science, there are a lot of *online communities*; they have generated such projects, trying to identify a *relationship from tags, events,* and *persons* connected by relations, and generate an ontology and *populate the ontology to the whole community*. Every day, more and more data is being generated. (From Group 2) I think that relationships in the person's section are potentially extremely useful when you want to find your way through this kind of historical data and *different relations between people*. But I am wondering, if you have John and *"wife of", "son of",* do you have a *"husband of", "father of" relation.*	Semantic Web Online communities The relationship between tags, events, person Community centered ontology Relationships between people are useful inverse relationship

Table 3.4 Sample emerging and superordinate themes

Step 4. Summarizing themes

The final step in the analysis produces a table summarizing the themes, superordinate themes, and quotes from participants to match each theme together. Each theme is considered individually, and at the same time, themes are closely interlinked.

3.3. Quality assurance of the research

Various steps were taken throughout the research process to ensure the methodological rigour of this research. Since a multi-method approach was chosen for this research, consequently, criteria pertaining to both quantitative and qualitative research methods were used to measure the quality of the research. Methods contributing to the validity, reliability, credibility transferability, and confirmability of the results of this research are explained below.

Validity addresses the “question of how well the … reality being measured through research matches with the constructs researcher use to understand it” (Neuman, 2003, p.179). Ensuring the validity of a study means identifying and addressing the potential methodological weaknesses that may also undermine the quality and eventual generalizability of the results obtained. To ensure the validity of this research, the three methods (i.e., content analysis, methontology, and focus groups studies) are cautiously combined and designed to meet the objectives of this research as well as ensure that what the studies intend to measure is measured (Neuman, 2003). As this research consists of Studies 1, 2, and 3, the validity of the study is directly related to the validity of each study phase as the three studies in the process are connected to one another. To be effective, the data collection instruments should directly link to the research questions and be applied consistently (Pickard, 2013; Widemuth, 2009). In order to do so, links between research questions and data collection instruments were established.

In the qualitative methodology, reliability refers to "whether the research process is consistent and is carried out with careful attention to the rules and conventions of qualitative methodology" (Ulin, Robinson, & Tolley, 2005, p. 26). High reliability of the research ensures that another research endeavor would be able to reproduce the same research and arrive at the same conclusions (Yin, 2003). In order to ensure reliability, in Study 1, a research protocol was developed for the research process. The data collection instruments were tested to ensure quality data was consistently collected. In particular, archival descriptions have been selected from two kinds of institution sites which are managed by different encoding schemes. Data sampling was done systematically through a combination of random sampling and purposive sampling to ensure that archival descriptions were selected from various institutions. In the content analysis of sample data in Study 1 stage 1 event identification, coding guidelines were based on a widely used guideline (Saurí, Littman, Gaizauskas, Setzer, & Pustejovsky, 2006). The coding guidelines were consistently applied to the data.

Based on the coding guidelines, a coding scheme was applied to the content analysis. These considerations could increase its dependability and confirmability.

In Study 2, since methontology is an engineering method for creating ontologies, each activity in the methontology procedure was carried out in this research as it has been defined. Reliable data sources (i.e., glossaries, terminology database) were selected by following detailed instructions. Inputs and outputs of each activity were identified. For example, in the specification stage, a clear specification was developed along with a list of competency questions and explicit objectives. Domain related sources have been utilized to develop specification requirements. In the conceptualization stage, various ontologies and event-related models (i.e., Simple Event Model, BFO, CIDOC) have been reviewed systematically. Reviewed ontologies and the results of the qualitative content analysis have been integrated into the developed ontology. The formalization stage further formalizes the model into an ontology based on description logic. The integration and implementation stages convert the model into RDF and OWL formats in further alignment with existing upper ontologies. The maintenance stage fixes any problem in the ontology and maintains the correctness of the ontology, both in syntax and semantics. In addition, *Protégé, which is a* "free, open-source ontology editor and framework for building intelligent systems" (Stanford Center for Biomedical Informatics Research, 2015), was utilized to create, edit, and verify the correctness of the ontology. The detailed procedures and clearly defined inputs and outputs of each stage of the methontology ensure that the method was consistently carried out.

Credibility refers to the "confidence that can be placed in the truth of the research finding" (Korstjens & Moser, 2018, p. 121). To increase the credibility of the research, multiple data sources have been used to improve the quality of the results. In Study 1, two lists of archival sites (each list consists of 20 archival sites) were selected. In Study 2, the results of Study 1, existing standards, and high-quality terminology databases were used. In Study 3, the design of the focus groups was deliberately constructed in an identical manner with modifications to each group. The researcher attempted to carry out both focus group sessions in an identical manner to one another as much as possible. The same format and content of the questions were used in the two sessions, with some questions modified to fit each group better. The moderator (i.e., the researcher) tried to encourage discussion of the addressed questions and pull out opinions from participants as much as possible. This design attempts to solicit broader opinions and minimize biases as much as possible. In this research, in order to ensure that data was consistently collected and analyzed, research tools (such as research protocols, codebook) were developed.

External validity and transferability are concerned with whether "findings from the investigation can be generalized to the wider context" (Pickard, 2013, p.22). This research aims to address the fundamental problem of archival description. The findings may not apply to all kinds of data in cultural heritage collections or digital humanities. However, the sample size of Study 1 achieved theoretical saturation, which means there is redundancy in the data collected from different archival sites. The results of Study 1 can be applied to archival descriptions produced in

North American settings. In Study 2, the results of Study 1 have parts of data sources that could affect the transferability of Study 2. However, other data sources in Study 2 are reliable and high-quality. With a rigorous Methontology process, the results of Study 2 can be applied to archival descriptions in North American settings. In Study 3, the sample size is too small to apply the results in a broader context.

To reveal the comparative results of the responses from the different groups, the design of the focus groups was comparatively constructed. Two kinds of participants were solicited (novices and professionals), the interview questions (see APPENDIX IV. Sample questionnaire) were modified to fit each group, two kinds of sample archival descriptions (see APPENDIX V. Three sample archival descriptions for focus group discussions) were provided to demonstrate their differences (e.g., existing archival descriptions and a proposed ontology model). In the data analysis, the transcripts were reviewed and double-checked by a Master's student who has experience with audio recordings and transcribing scripts. A thematic analysis of the focus group was carefully applied to the study.

Confirmability refers to "the extent to which the characteristics of the data can be confirmed by others who read or review the research results" (Bradley, 1993, p. 437). The objectives, research questions, and corresponding protocols are explicitly defined. Confirmability is "vital to limit investigator bias" (Pickard, 2013, p. 22). Such bias in research includes "the researcher's worldview, disciplinary assumptions, theoretical proclivities, and research interests" (Charmaz, 1995, p. 32). The conformability of the results was ensured by the sets of procedures developed for data collection and analysis in the three studies. The explicitly defined procedures and codebooks make the reproduction of the study by other researchers possible. The results can be traced back to the raw data of the research. In the research, multiple data sources (i.e., two sets of archival sites in Study 1, multiple terminology databases, glossaries in Study 2, and two groups in Study 3) were used to triangulate and minimize biases from data sources.

3.4. Limitations of the study

Due to the relatively small sample size of archival descriptions in this study, generalizations beyond the context of this study may not be easy to infer. The data was first collected using content analysis and subsequently, two focus groups were conducted with only six users and five archival professionals. The results of this study may not represent the entire population of users and archival professionals. In addition, this study focused on archival descriptions that were created in the United States of America and Canada only; thus, the results may not reflect the archives in countries beyond North America.

3.5. Ethics

In order to elicit users' opinions on Event Ontology for Archival Context, a study with two focus groups was carried out after receiving ethics approvals from the McGill University Research Ethics Board Office (REB-II) (see APPENDIX VII. McGill University Research Ethics Board approval form) and the Lakehead University Research Ethics Board Office (see APPENDIX VIII. Lakehead University Research Ethics Board approval letter). In general, this study was carried out with respect to human dignity and will not expose the participants to any foreseeable harm or risks.

Summary

This chapter has described the methodological components of this research. This study is based on a multi-method approach with three studies: (1) content analysis; (2) methontology; and (3) focus group. In the content analysis (Study 1), archival descriptions collected from two sets of archival sites were analyzed. In Study 2, the procedures of methontology were followed. In Study

3, discussions collected from two focus groups were coded and analyzed. In the three studies, various means have been taken by the researcher to ensure the quality of research in terms of its preparation (such as the research protocol, pre-test, and use of multiple data sources), and the research was carried out as planned.

Based on the designed methods, the next chapter explains in detail the findings of these Studies.

CHAPTER 4. THE FINDINGS OF STUDIES 1 & 2

This chapter presents the findings of Study 1 and Study 2 in two sections. The first section presents the results of the content analysis (Study 1). The second section explains the findings of the methontology (Study 2).

4.1. Entities and relationships in archival descriptions

The first section presents the findings to answer research question (RQ) 1 (which entities are the essential entities in archival descriptions, especially archival context?) and two sub-questions (RQ 1.1: which entities are used to provide information on provenance and context in existing archival descriptions and RQ 1.2: what are the relationships between these entities when describing archival context?).

A content analysis (Study 1) was conducted to answer RQ1 by identifying the entities of archival descriptions and examining their specific relationships in terms of archival contexts. To identify entities and their relationships, the following four-step procedure was conducted:

1) Identifying general entities in archival descriptions
2) Identifying sentences in archival descriptions
3) Identifying granular entities in archival descriptions
4) Categorizing events in archival descriptions

Each step in the procedure is explained in the following sections.

4.1.1. Identifying general entities in archival descriptions

Archival descriptions, especially archival contexts, contain information related to the creator(s) of archival materials, relevant people, organizations, and societal backgrounds. Here, the entity could be considered a container that comprises many sentences. Each sentence consists of smaller entities, including events. Entities are simply grouped by the original elements (or tags) that are assigned by the individual who created the archival descriptions. In this sense, from general to specific, archival descriptions are categorized into entities, sentences, and events. Descriptive statistics can reveal the general composition of entities in archival descriptions. Thus, archival descriptions, especially archival contexts, consist of several entities, such as abstract, administrative history, biographical information, custody history, other notes, related information, and scope & content. Entities are also called elements in EAD (Encoded Archival Description) version 2002 and EAD 3.

Regarding the two sets of sample archival descriptions that have been collected for this research, while Set 1 was retrieved and collected from the Repositories of Primary Sources (Abraham, 2015) site, Set 2 was from the SAA EAD implementors' sites (Society of American Archivists, 2010). There are 120 archival descriptions at the fonds level retrieved from 40 archival institutions sites, as explained in Chapter 3, section 3.1. Each description has been further dissected into general entities by abstract, administrative history, biographical information, custodial history, notes, and scope & content. Set 1 consists of 231 entities and Set 2 has 288 entities, respectively. In total, 519 entities have been dissected. In Set 1, 231 entities were identified as four kinds of general entities: administrative history, custody history, biographical information, and scope & content. In Set 2, 288 entities were identified by seven kinds of general entities: abstract, administrative history, biographical information, custody history, other notes, related information, and scope & content. Figure 4.1 shows the percentage distribution of general entities in both sets of the sample archival descriptions.

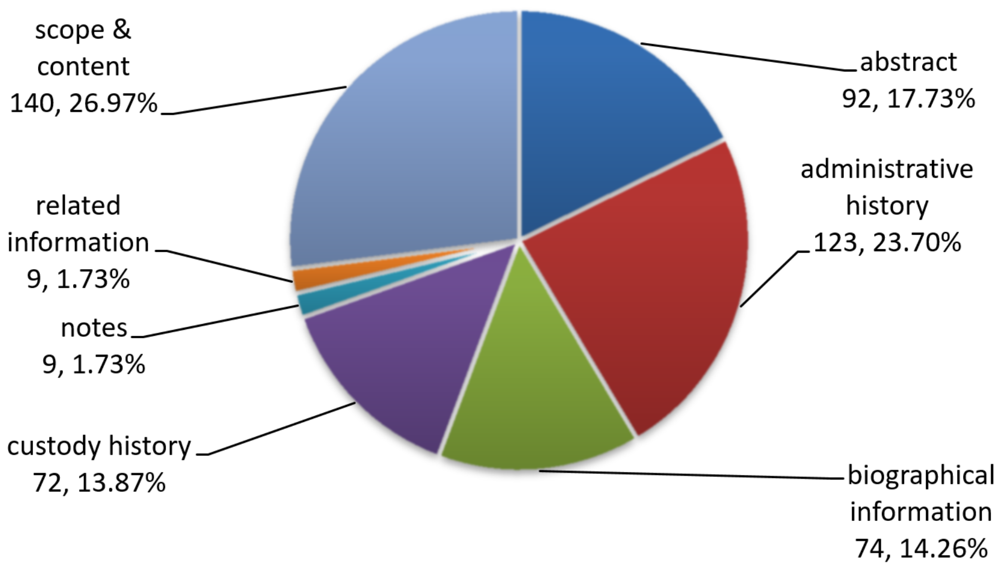


Figure 4.1 Distribution of general entities in sets 1 and 2.

Among the 519 entities, five kinds of general entities cover 96.53% of archival descriptions in the following order: scope & content (140, 26.97%), administrative history (123, 23.70%), abstract (92, 17.73%), biographical information (74,14.26%), and custody history (72, 13.87%).

4.1.2. Identifying sentences in archival descriptions

A sentence is a linguistic element and refers to "one or more main clauses, corresponding to units in which in written language are bounded by the punctuation mark" (Preisler, 1997, p. 24). To identify sentences, punctuation marks, especially the period mark ("."), are used. Although a period is used in some cases (e.g., Dr., Jr., etc.), computer algorithms can achieve high accuracy with over 95% precision and 90% recall to detect sentences in natural language application programs (Agarwal, Ford & Shneider, 2005) or they can be manually identified. In this research, sentences are automatically identified and then manually reviewed. In total, 3,603 sentences are identified in Sets 1 and 2. Figure 4.2 shows the percentage distribution of entities by the number of sentences in both sets.

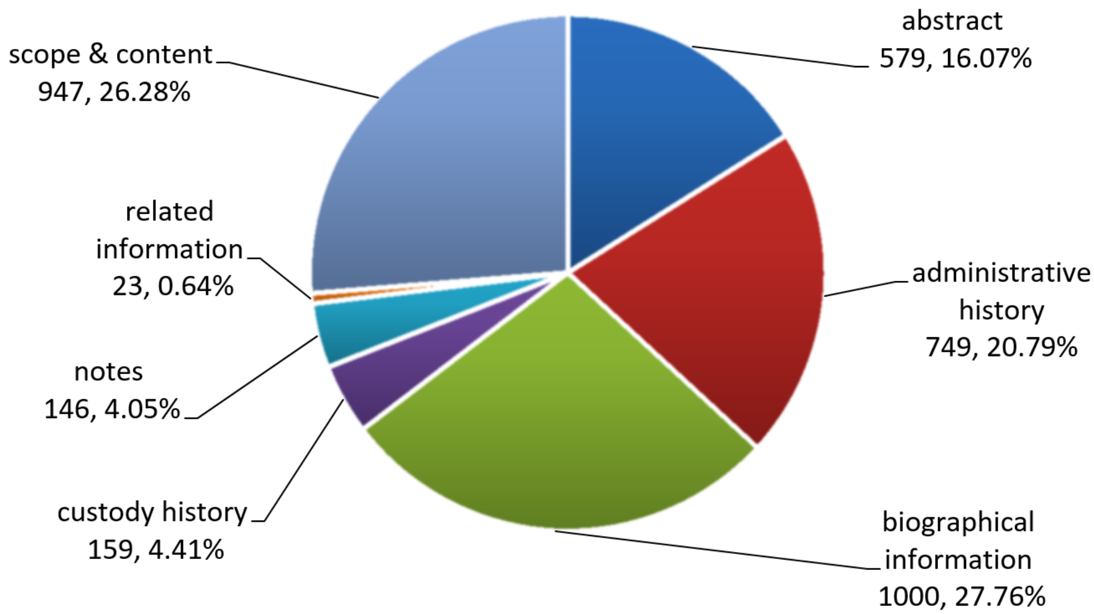


Figure 4.2 Distribution of sentences by entity in sets 1 and 2

One general entity contains one or more sentences. For example, the number of sentences in the custody history entity is 159 sentences out of 3,603 sentences (4.41%) illustrated in Figure 4.2, while the number of the custody history entity is 72 entities out of 519 entities (13.87%) shown in Figure 4.1. In comparison with biographical information, the number of sentences in biographical information is 1,000 out of 3,603 (27.76%) and the number of entities in the biographical information is 74 out of 519 entities (14.26%). This indicates that the distribution of sentences and entities are not even in archival descriptions. The next step is to take a closer look at identifying entities.

4.1.3. Identifying granular entities in archival descriptions

RQ 1.1 identifies entities in archival descriptions, especially in archival contexts. In general, an entity is defined as "a physical, digital, conceptual, or another kind of thing with some fixed aspects" (Moreau & Groth, 2013, p. 24). In a narrow sense, entities refer to "all objects or agents of the system being analyzed" (Pearce-Moses, 2005, p. 148). Relations refer to linkages between entities and can be treated as a special type of entity. General entities (e.g., scope & content, abstract, administrative history, custody history, related information, biographical information, and notes) of archival descriptions are widely used in describing archival collections in archives, as explained in the previous section. General entities consist of granular entities and relations.

In each archival description, the coding process was iterative with existing general entities when coding the 120 archival descriptions. After the coding, while the general entities remain valid, more granular entities emerged, such as *Agent*, *Artifact*, *Event*, *Feature*, *Function*, *Place*, *Relation*, *Role*, *State/Situation*, and *Time*, as summarized in Table 4.1.

Granular Entity	**Description**
Agent	The person (s), corporate body(ies), or family(ies) chiefly responsible for the intellectual or artistic content of a document (Bureau of Canadian Archivists - Planning Committee on Descriptive Standards, "APPENDIX D: Glossary," Rules for Archival Description (Ottawa: Bureau of Canadian Archivists, July 2008, p. D-2)
Artifact	Something that is deliberately designed by agents to address a particular purpose (Basic Formal Ontology, 2015).
Event	Something occurs over a period of time at certain places and acts upon or with entities (Pustejovsky et al., 2005).
Feature	“Properties or qualities as distinguished from any particular embodiment of the properties/qualities in a physical medium” (Standard Upper Merged Ontology, 2012, para. 1)
Function	The “activities of an organization or individual performed to accomplish some mandate or mission” (Pearce-Moses, 2005, p. 179).
Place	“A designated point or region in space” (Lee, 2011, p. 112)
Relation	An association between two or more entities to connect entities
Role	Something that agents carry out because of certain external circumstances (Basic Formal Ontology, 2015)
Situation	An entity that is associated with a period of time where a set of statements is true (Lagoze & Hunter, 2002).
Time	A temporal entity with zero extents or duration or an extent or duration (Hobbs & Pan, 2005)

Table 4.1 Identification of granular entities in archival descriptions

Each granular entity is explained in the next section.

4.1.3.1. **Agent**

Agents are authors who are the "persons, corporate bodies, or families chiefly responsible for the intellectual or artistic content of a document" (Bureau of Canadian Archivists - Planning Committee on Descriptive Standards, 2008, p. D-2). For instance, "James L. G. Fitz Patrick" is the author of his research publications. Agents have "a role in the lifecycle of a resource" (Wilson & Clayphan, 2004, Para 2. Section 2) and are directly or indirectly related to an information object. For example, "Peter Spiridon" and "Harold J. Smith" from Table 4.1 may not directly relate the creation of the archival material of the fonds. However, they were directly related to "James L. G. Fitz Patrick". "The College of Staten Island" and "Staten Island Community College", and "the College Archives" provide contextual information that may potentially link to other individuals or organizations associated with the three corporate bodies. Creators are often related to other persons within and out of their family and organizations. Other persons or organizations may not be directly related to an information object, but they may have influence on the person who created the record. That is why persons, corporate bodies, and families in archival descriptions are important agents.

To explain each agent, the following examples are taken from the sample archival descriptions. [ ] is used with a type of agent to represent in the examples:

1) Person

Sentence 1: *Charles Brockden Brown* [Person] is considered the first American novelist, although he also wrote short stories, essays, and political pamphlets.

Sentence 11: During this period, *Brown* [Person] met *William Dunlap* [Person] and other members of the New York literary society the Friendly Club through *Elihu Hubbard Smith* [Person] Brown's "closest friend and patron" (Chapman).

(Source: Set 1, Institution #9, Archival Description #1)

In Sentence 1, Charles Brockden Brown is identified as a creator. In Sentence 11, William Dunlap and Elihu Hubbard Smith are two individuals who played roles in Brown's life. They are not direct creators of Brown's works but may have had an influence on Brown.

2) Corporate body

Sentence 1: The four room *Bridgeport School* [Corporate body] was constructed in 1908 although only one room on the first floor was occupied for the first few years.

(Source: Set 1, Institution #5, Archival Description #1)

In Sentence 1, Bridgeport School is a corporate body creator identified in the description.

Sentence 1: *Cooper & Beatty, Limited* [Corporate body] was founded in 1921 by *E. Cooper* [Person], *L. Beatty* [Person] and *J.L. Pepper* [Person] using the name *Trade Composition Company* [Corporate body].

(Source: Set 1, Institution #13, Archival Description #3)

This sentence contains two corporate bodies and three individuals.

3) Family

A family consists of many members and the history of a family is a collective history of the members. The following example description includes the members of the family and their relationships with each other.

Sentence 1: *Rev. Alfred Chipman* [Person] was born in 1834 at Pleasant Valley (now Berwick), Nova Scotia, the son of *Rev. William Chipman* [Person] and his second wife *Eliza Ann* [Person].

Sentence 4: *Alice Theodosia Shaw* [Person] was born 21 June 1832, the daughter of *Isaiah Shaw* [Person] and *Sarah Lyons* [Person].

Sentence 16: Alfred and Alice were married at River Philip, Nova Scotia on 28 October 1862.

Sentence 17: They had three children: *Fred McCullock* [Person] who was born in 1865 at Pictou, Nova Scotia and who died in Florida in 1892; *Alvah Hovey* [Person] who was born in 1867 in Stewiake, Nova Scotia (donor of these records); and *Chaloner Oaken* [Person], born 1871 at Sydney, Nova Scotia.

Sentence 20: *Isaac Chipman* [Person] was a half-brother of Rev. Alfred Chipman and was born at Cornwallis, Nova Scotia on 17 July 1817 son of Rev. William Chipman and his first wife.

(Source: Set 1, Institution #1, Archival Description #2)

Based on the description above, a family tree can be manually drawn from the identified relations. If the text was semantically marked (such as person, relation), the family tree could be automatically generated. This illustrates information embedded in the text that could be identified and presented differently.

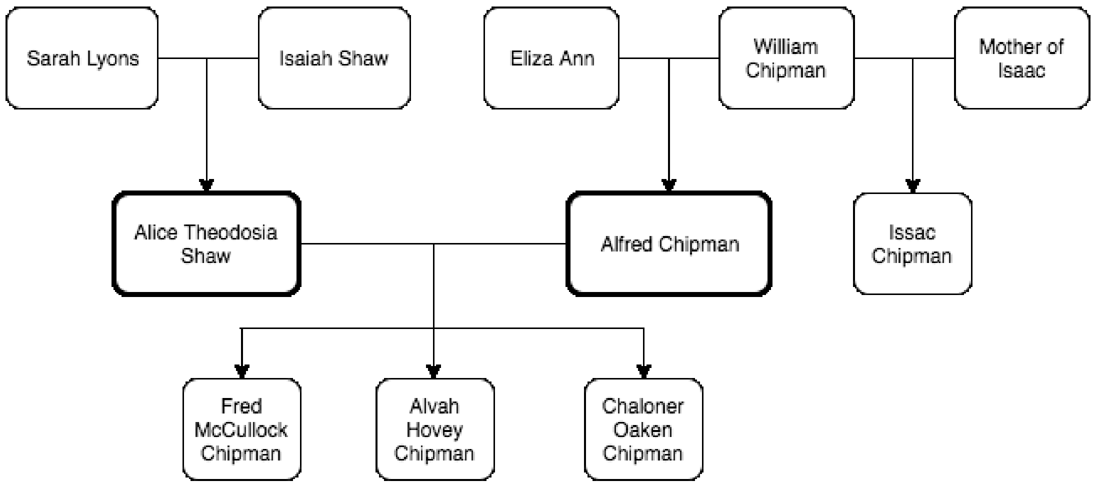


Figure 4.3 A family tree based on the agent's relations

This family tree represents how two persons, *Alfred Chipman,* and *Alice Theodosia Shaw,* are directly related to each other and their other family members.

4.1.3.2. **Artifact**

Artifacts indicate something that is "deliberately designed by agents to address a particular purpose (Basic Formal Ontology, 2015)." It is "deliberately created … by human beings to be used to achieve certain goals. The[ir] essential features have to do with the purpose or use for which the artifact was created" (Arp, Smith, & Spear, 2015, p. 70). As a type of artifact, an information artifact is "a subtype of BFO's "generically dependent continuant" (Arp et al., 2015, p. 168). In this study, artifacts are considered records (e.g., information artifacts) created by creators; they can be any physical object (e.g., artifacts) that plays a role in the creation of records (e.g., information artifacts). For example, "news clippings" and "research publications" from Table 4.1 are information artifacts. News clippings can be categorized as “non-scholarly publications” and research publications as “scholarly publications”.

Among the artifacts identified in this study, the focus of this step was to identify whether they are information artifacts or other types of artifacts since information artifacts can be created as records or related artifacts included in the archival fonds. Information artifacts identified in this study are listed in APPENDIX IX. Information artifacts.

4.1.3.3. **Event**

An event refers to something that occurs over a period of time at certain places and acts upon or with entities (Pustejovsky et al., 2003). For example, in Table 3.3, four events have been identified. The criteria to identify and categorize events and the procedure to code the identified events are explained in section 4.1.4. Categorizing events in archival descriptions.

4.1.3.4. **Feature**

A feature is a prominent attribute or aspect of something. Features provide valuable information about the object that is described in archival descriptions. For example, in the following sentence:

> Sentence 1: The four room *Bridgeport School* [Corporate body] was constructed in 1908 although only one room on the first floor was occupied for the first few years.
> (Source: Set 1, Institution #5, Archival Description #1)

“The four room” can be considered a feature of the Bridgeport School, revealing that the school has only four rooms. Depending on the physical medium, the number and types of features can vary.

4.1.3.5. **Function**

Functions refer to the "activities of an organization or individual performed to accomplish some mandate or mission" (Pearce-Moses, 2005, p. 179). The activities of an organization are carried out by individuals who are members of the organization. Individuals play certain role(s) in carrying out a specific function of an organization. One individual may have more than one role in the realization of a function. An organization may have multiple functions. For example, a student union is a student organization that carries out certain functions:

> Sentence 4: Defined by the university as a student organization, the union has an Activities Board, made up of students, which *plans* [function] and *coordinates programs* [function].
> (Source: Set 2, Institution #16, Archival Description #3)

In the example above, the functions of the student union are to plan and coordinate programs.

4.1.3.6. **Place**

A place is a "designated point or region in space" (Lee, 2011, p. 112) as a spatiotemporal region. Place information can be "particularly informative when associated with occurrences and agents" (Lee, 2011, p. 112). Time and place provide important background information to creators' activities into context.

In archival descriptions, the name of a place can be a city, region, town, or street. Here are some examples identified in the study:

1) Region

> Sentence 1: The collection was donated in 1996 by Dr. Fred C. Leone of *Silver Spring,* Maryland [place].
> (Source: Set 2, Institution #1, Archival Description #1)

2) City

> Sentence 1: The Right Reverend Peter Guilday was born on March 25, 1884 in *Chester, Pennsylvania* [place].
> (Source: Set 2, Institution #1, Archival Description #2)

3) Street

Sentence 1: Banks & Co Ltd, a firm of engravers and printers, was situated at *12 George Street, Edinburgh, Scotland* [place], by at least 1929, they later expanded to occupy *53 George Street* [place], this taking place by 1968.

(Source: Set 2, Institution #1, Archival Description #2)

4) Province

Sentence 6: He retired from business in the first decade of the 20th century, spending his summers in *Nova Scotia* [place] and his winters in *Boston* [place] and *New York* [place].

Sentence 7: He died in *Nova Scotia* [place] in October 1933.
(Source: Set 1, Institution #10, Archival Description #2)

Like the examples above, different types of places were identified and marked as places. As the concept of place is a spatiotemporal concept, it is bound to time, and they are often together. For example, at a certain period (before 1970), the red river road was located in the city of Port Arthur, Ontario. After 1970, the red river road was located in the city of Thunder Bay. The geographical location of the road remains in the same, but the name of the place can be changed.

4.1.3.7. **Relation**

A relation is an association between two or more entities (i.e., Concepts and Event) to connect entities. Relations can be categorized into two types: universal and particular (Arp et al., 2015). A universal is "an entity with a spatiotemporal existence … at any time" (Neuhaus, Grenon, & Smith, 2004, p. 49) and universals exist when and where they are instantiated as instances (particulars). A particular is "an instance of one or more universals" (Neuhaus, Grenon, & Smith, 2004, p.49). A relation can exist between two universal or particular entities, one universal entity and one particular entity. For example, Bill Clinton (a particular) is a *man* because he instantiates the universal *man*. Relations held between two universals ("universal-universal"), between a particular and a universal ("particular-universal"), and between two particulars ("particular-

particular") (Arp et al., 2015, p. 133). In the three major families of relations, the common relation in universal-universal is *is-a relation,* such as a person is a type of agent; *instance_of* is a common relation between particular and universal (e.g., Bill Clinton is an instance of man); part_of is between two particulars (e.g., this tire is part of this car).

In applying the relation to the context of archival descriptions, relations exist between the entities identified in archival descriptions. The entities are instances of their classes. In other words, the relations identified are at the instance level.

The following example shows different types of relations:

1) Relations between two and more agents.

Sentence 20: *Isaac Chipman* [Agent] was a *half-brother* [Relation] of *Rev. Alfred Chipman* [Agent] and was born at Cornwallis, Nova Scotia on 17 July 1817 *son* [Relation] of *Rev. William Chipman* [Agent] and *his first wife* [Agent].
(Source: Set 1, Institution #1, Archival Description #2)

Half-brother is a relation between Issac Chipman (who is an instance of a person and person is an agent) and Alfred Chipman (who is another instance of person). Issac Chipman is also the son of William Chipman and his first wife.

2) The relation between an agent and an organization.

Sentence 1: *Sandor Teszler* [Agent] was a long-time *member* [Relation] of *the Wofford Community* [Organization].
(Source: Set #1, Institution #19, Archival Description #4)

A member-of relation can be identified between Sandor Teszler, who is an instance of person, and person is an agent and the Wofford Community, which is an instance of an organization.

3) The relation between two organizations.

Sentence 11: During the first half of the 20th century *Acadia University*[organization] was a member of *the Maritime Provinces Branch of the Amateur Athletic Union of Canada (MPBAAUC)*[organization].
(Source: Set 1, Institution #1, Archival Description #1)

A member-of type of relation is identified in the example above.

Identified relations are categorized into three broad entities: 1) kinship relation (e.g., *son-of*, *daughter-of*, *parent-of*, etc.; 2) social relation (e.g., *colleague-of*, *friend-of*, etc.); and 3) other relations to link identified entities, except for the generic *is-a, part-of,* and *is-instance-of* (e.g., an event *has_participant* John, an event *occur_atTime* year 2005, an event *occur_atPlace* city of Toronto, etc.)

4.1.3.8. **Role**

A role is a "realizable entity that is possessed by its bearer because of some external circumstances" (Arp et al., 2015, p.99). A realizable entity is defined as "a specifically dependent continuant that has at least one independent continuant entity as its bearer, and whose instances can be realized (e.g., manifested, actualized, executed) in associated processes of specific correlated types in which the bearer participates" (Arp et al., 2015, p.98). So, agents play many different roles. In some cases, agents do not always have to be in the given external circumstances, and the agent can have functions, features, and situations. For example, the role of an author of a document is realized in the process of creating the document.

In this study, several types of roles were identified (e.g., community role, professional role, or military role). For example, see the following sentences:

Sentence 5: Born in Orebro, Sweden, Edita deToll Morris (1902-1988) began her *career writing short stories* [professional role].
Sentence 8: In 1957, *the couple founded Hiroshima House* [community role], a recreation center and hostel for the survivors of the atomic bombings.
Sentence 9: In addition, the Morris's were *political activists* [political role] committed to nuclear disarmament, opposed to many U.S. policies of the Cold War, and related causes.
(Source: Set 2, Institution #5, Archival Description #1)

In the example, “Edita deToll Morris” (an agent) has played at least three different roles: professional role (as an author), community role (as a founder), and political role (as an activist).

4.1.3.9. **Situation**

A situation is, loosely, an entity that is associated with a period where a set of statements is true (Lagoze & Hunter, 2002). During the coding process, sentences containing “being at a place”, “being in a relationship”, “having a lung cancer”, and “experience illness” were identified as situations of specific agents. The selected phrases in some sentences could be considered the results of an event. For example, “being at a place” may be the result of a travel event and “being in a relationship” is the result of a “beginning a relationship” event. In comparison, "having a lung cancer" and "experience illness" present statuses or states of entities. In these two examples, they denoted that an agent had lung cancer and an agent was ill. They are time-bound situations, meaning that the situations may not exist before the agent was ill and after the agent was recovered or died. In this study, medical, political, financial, and military situations are commonly identified in the sample archival descriptions.

4.1.3.10. **Time**

Time is a spatiotemporal entity, as a critical aspect of both provenance and context. Time can be categorized as instant and interval. Intervals are "things with extent and instants are … point-like in that they have no interior points" (Hobbs & Pan, 2006, Section 2 para. 2). Instant is a point in time and can be considered an interval with the same at the beginning and end.

In archival descriptions, time can be easily identified as:

1) Small Interval

Sentence 18: Alice and Alfred retired to Berwick, NS in *1899* [time].
Sentence 19: Alfred died *24 April 1918* [time], at Berwick, while Alice died at the age of 88, on *18 January 1921* [time], also at Berwick.

Sentence 20: Isaac Chipman was a half-brother of Rev. Alfred Chipman and was born at Cornwallis, Nova Scotia on *17 July 1817* [time] son of Rev. William Chipman and his first wife.
(Source: Set 1, Institution #1, Archival Description #2)

2) Interval

Sentence 11: *During the first half of the 20th century* [time] Acadia University was a member of the Maritime Provinces Branch of the Amateur Athletic Union of Canada (MPBAAUC).
(Source: Set 1, Institution #1, Archival Description #1)

As the identification and application of time often involve imprecision and levels of granularity (Allen, 1983), this can be observed in time-related expressions in the descriptions. For example, the granularity level of "24 April 1918" in the sentence 19 is a day, which is smaller than that of "1899" in sentence 18. For an interval, it may span over a year, a month, and a half-century like an interval identified in sentence 11.

4.1.4. Categorizing events in archival descriptions

In general, an event is defined as something that occurs over a period of time at certain places and acts upon or with entities (Katz, Pustejovsky, & Schilder, 2005). In coding archival descriptions, more specifically, an event is considered “a special occurrence involving participants”, “something that happens”, or “described as a change of state” (Pustejovsky et al., 2003, p.53). Since events are “most typically represented by verbs … also possible for them to be represented by nouns” (Styler IV, et al., 2014, p. 145), thus, in this study, only two types of events are identified in the archival descriptions.

The largely manual procedure to identify events is similar to that of identifying granular entities and assigning them to event types. To explain the procedure for identifying events, the previous procedure for identifying granular entities can be extended with one more step to assign appropriate event types as follows:

1) Identifying and locating events described by verbs and nominalized verbs
2) Identifying other phrases associated with each verb
3) Marking appropriate spans of text to represent the phrases
4) Assigning appropriate entities to the phrases
5) Assigning appropriate event types to the events and linking events with related entities

In the following example, four events have been identified, including three donation events as verbs and one retirement event as a noun.

> Sentence 1: The vast majority of the collection *was donated to* the College of Staten Island by James L. G. Fitz Patrick's stepson, Harold J. Smith, in 2000.
>
> Sentence 2: A small number of news clippings *were donated by* Peter Spiridon, a former colleague at Staten Island Community College.
>
> Sentence 3: James L. G. Fitz Patrick also *donated* several of his research publications to the College Archives prior to his *retirement* in 1976.

From Steps 1 to 4 above, events were identified from each sentence by four types of entities in Set 1 (i.e., administrative history, biographical information, custody history, and scope & content) and seven types of entities in Set 2 (i.e., abstract, administrative history, biographical information, custody history, notes, related information, and scope & content). To see the distributions of the three percentage changes (along with Table 4.1 and 4.2), Table 4.2 demonstrates the comparative summary of entities, sentences, and events in Sets 1 and 2.

Entity	**Set 1**			**Set 2**		
	Entity	**Sentence**	**Event**	**Entity**	**Sentence**	**Event**
abstract				92 (31.94%)	579 (28.35%)	1,389 (28.65%)
administrative history	67 (29.52%)	475 (30.43%)	1,139 (32.08%)	56 (19.44%)	274 (13.42%)	434 (8.95%)
biographical information	43 (18.94%)	444 (28.44%)	1,006 (31.15%)	31 (10.76%)	556 (27.23%)	1,517 (31.28%)
custody history	27 (11.89%)	56 (3.59%)	119 (3.35%)	45 (15.63%)	103 (5.04%)	138 (2.85%)

notes				9 (3.13%)	146 (7.15%)	442 (9.12%)
related information				9 (3.13%)	23 (1.13%)	22 (0.45%)
scope & content	90 (39.65%)	586 (37.54%)	1,286 (36.22%)	46 (15.97%)	361 (17.68%)	907 (18.70%)
Total	227 (100%)	1,561 (100%)	3,550 (100%)	288 (100%)	2,042 (100%)	4,849 (100%)

Table 4.2 Summary distributions of entities, sentences, and events in sets 1 & 2

In Set 2, the percentage of custody history has decreased from 15.63% (entity) to 5.04% (sentence) and finally to 2.85% (event). The percentage of related information has also decreased from 3.13% (entity) to 1.13% (sentence) and to 0.45% (event). The biographical information of Set 2 has a lower number and percentage of entity (10.76%) in comparison to those of sentences and events. This may be caused by the archival description samples of Set 2, which are in EAD forms. In the meantime, the percentage of abstract and scope content shows a relatively even distribution over entity, sentence and event within each set. The data shows that information in archival descriptions are unevenly distributed among entities. The number of events is greater than the number of entities or sentences.

In the procedure to identify events, the final step is to assign appropriate event types to the identified events. TimeML is a comprehensive and "robust specification language for events and temporal expressions in natural language" (Pustejovsky et al., 2003, p. 28) so that TimeML is selected to code the events in this study. In TimeML, there are seven types of events that can be assigned to the identified events. In the TimeML event guideline, event is categorized into seven types of sub-classes or sub-entities by their features or states, including Aspectual, I_Action, I_State, Occurrence, Perception, Reporting and State (Saurí et al., 2009). These types of events are defined and described in detail in Table 4.3 (summarized from Saurí et al., 2009).

Type of events	Definitions	Examples
Aspectual	An event consists of an aspectual "predicate which selects an event as its argument, and points to some structural aspect of the event" (Saurí et al., 2009, p. 16)	The volcano *began* showing signs of activity in April for the first time in 600 years (Saurí et al., 2009, p. 17). **Verbs**: begin, finish, stop, continue
I_Action	Standing for "intentional action"; "dynamic events for an event-denoting argument, which must be explicitly present in the text" (Saurí et al., 2009, p.17).	Palestinian police *prevented* a planned pro-Iraq rally by the Palestinian Professionals' Union (Saurí et al., 2009, p. 18). **Verbs**: attempt, try, prevent, promise, offer
I_State	Standing for "intentional state"; events that "select an argument that expresses any sort of event" (Saurí et al., 2009, p.18).	We *believe* that his words cannot distract the world from the facts of Iraqi aggression (Saurí et al., 2009, p. 18). **Verbs**: believe, intend, want
Occurrence	"all of the many other kinds of events that describe something that happens or occurs in the world" (Saurí et al., 2009, p.20).	16 planes have *landed* so far with protective equipment against biological and chemical warfare (Saurí et al., 2009, p. 18). **Verbs**: crash, build, land, merge, sell
Perception	"the physical perception of another event" (Saurí et al., 2009, p.16).	Witnesses tell Birmingham police they *saw* a man running (Saurí et al., 2009, p. 16). **Verbs**: see, hear, watch, feel
Reporting	"the action of a person or an organization declaring something, narrating an event, informing the addressee(s) about a situation, and so on" (Saurí et al., 2009, p. 16).	Punongbayan *said* that the 4,795-foot-high volcano was spewing gases up to 1,800 degrees (Saurí et al., 2009, p. 16). **Verbs**: say, report, announce
State	"States describe circumstances in which something obtains or holds true" (Saurí et al., 2009, p. 19)	Iraq *has* 17 million residents (Saurí et al., 2009, p. 19). **Verbs**: have kidnapped, love

Table 4.3 Seven types of events

These categorizations of events are used to group events (Derczynski, 2017). For instance, in the examples of Sentences 1-3 above, *donation* events fall into the type of "Occurrence" and *retirement* event is an "Aspectual" type of event. Table 4.4 displays the number of events categorized by the seven types of events in two Sets.

	Aspectual	I_Action	I_State	Occurrence	Perception	Reporting	State	Total

Set 1	Administrative History	29	11	13	1,055	0	1	24	1,139
	Custody History	0	0	3	109	0	1	6	119
	Biographical Information	35	12	6	930	2	3	18	1,006
	Scope & Content	11	5	5	1,237	0	2	17	1,285
	Sub-Total	75 (2.11%)	28 (0.79%)	27 (0.76%)	3,331 (93.86%)	2 (0.06%)	7 (0.20%)	65 (1.83%)	3,549 (100%)
Set 2	Abstract	23	10	4	1,307	5	6	34	1,389
	Administrative History	3	3	2	398	2	0	26	434
	Biographical Information	42	15	36	1,359	7	15	43	1,517
	Custody History	0	1	0	125	1	0	11	138
	Notes	19	4	9	386	3	6	15	442
	Related	1	0	0	20	0	0	0	22
	Scope & content	11	4	3	857	2	3	27	907
	Sub-Total	98 (2.02%)	37 (0.76%)	54(1.11%)	4,452 (91.81%)	21 (0.43%)	30 (0.62%)	157 (3.24%)	4,849 (100%)

Table 4.4 Number of events by type in sets 1 and 2

The majority of events are the “occurrence” type event in both Sets 1 and 2 in Table 4.4 Number of events by type in sets 1 and 2. While Set 1 has, among the identified 3,549 events, 3,331 events (93.86%) as occurrences, Set 2 has 4,452 events (91.81%) out of total 4,849 as “occurrence” type. Together, in both sets, 7,783 events (92.68%) are the occurrence type of events, which reveals that the occurrence type of events takes the majority of events, while the rest of events are fewer in number. The other six types of events are not evenly distributed in the sample archival descriptions. The occurrence type events seem to provide too general information about the events and archival descriptions consist of non-fictional and factual narrations of archival records in archives as so-called “neutral repositories of facts” (Schwartz & Cook, 2002, p. 1).

In step 5, in order to assign a proper and meaningful type of event to the identified event, it is necessary to take a closer look at the events. Events can vary depending on their context in each archival description and can be grouped into sub-classes or sub-entities. In the content analysis, as an initial step, general events are further thematically divided into granular events. A part of the identified granular events is displayed in Table 4.5. The entire list of granular events is listed in APPENDIX X. A list of events.

The total number of the identified granular entities is 2,645. These granular events can be further divided into smaller events. For example, the BioEvent can be a BirthEvent or DeathEvent. More granular events can be assigned and associated with involved agents. The following example shows how events are identified and assigned. In the example, there are six sentences. The identified events are marked with events in [ ].

Sentence 19: Alfred *died* [BioEvent] 24 April 1918, at Berwick, while Alice *died* [BioEvent] at the age of 88, on 18 January 1921, also at Berwick.
Sentence 20: Isaac Chipman was a half-brother of Rev. Alfred Chipman and was *born* [BioEvent] at Cornwallis, Nova Scotia on 17 July 1817 son of Rev. William Chipman and his first wife.
…
Sentence 25: He had taught [EducationEvent] for twelve years when, during a geology field trip [TravelEvent] to Cape Blomidon on 7 June 1852, Chipman and several others were *drowned* [BioEvent] when a storm came up and swamped [NaturalEvent] the boat in which they were returning [TravelEvent] to Wolfville.
Sentence 26: Chipman was *buried* [LifeEvent] in the Old Burying Ground, Wolfville.
(Source: Set 1, Institution #1, Archival Description #2)

Events	Description
ArchivalEvent	Events are concerned with archival related activities.
BioEvent	Events are concerned with an individual's birth and death.
BusinessEvent	Events are related to business transactions (e.g., ownership)
CommunicationEvent	Events are related to communication.
EducationEvent	Events are concerned with an individual who plays a role related to his/her education.
LegalEvent	Events are concerned with legal related activities.
LifeEvent	Events of an individual other than birth and death
MilitaryEvent	Events are concerned with military-related activities.
NaturalEvent	Events that happened naturally without a direct actor in the event (e.g., earthquakes and volcanic eruptions)
OrganizationalEvent	Events are concerned with an organization or organizational changes.
OtherSocialEvent	Other types of social events not included above
PoliticalEvent	Events are related to political activities
TravelEvent	Events are concerned with a move from one location to another

Table 4.5 An excerpt of descriptions of 13 types of events

In assigning granular events, the types of events are not mutually exclusive. *BioEvent* may be considered a special class of *LifeEvent*.

The initial classes of events tend to be partially related to a specific domain to capture their basic characteristics. For example, *LegalEvent* deals with events related to the legal domain. The initial classes of events are at a high level and further identified or merged. For example, *OrganizationalEvent* can be divided into the following sub-classes of events: *BeginAnOrganizaiton*, *OrganizationMerge*, *EndAnOrganization*, and *RenameAnOrganization*. These four sub-classes of events are the most popular events describing organizational changes.

In summary, through the content analysis (Study 1), sample archival descriptions have been analyzed. The characteristics of archival descriptions are presented by general types, sentences, and types of events. The procedures of five steps have been conducted to identify entities and relationships in archival descriptions. Through identification of entities, sentences, and events in archival descriptions, categories of entities emerged. As a result, ten types of contextual entities have been identified, including agent, artifact, event, feature, function, place, relation, role, situation and time. In addition, 13 types of events (2,645) are identified.

4.2. Study 2 - ontology development and representation of archival descriptions

Based on the findings of Study 1, this section describes the detailed procedures to answer Research Question 2 in Study 2 (i.e., to what extent can ontology technologies be used for archival context?). As explained in Chapter 3, methontology is a systematic approach to developing ontologies from the beginning. Methontology consists of three groups of distinctive activities: management, development, and support activities. Management activities include schedule, control, and quality assurance activities (Corcho et al., 2005). Development activities consist of specification, conceptualization, formalization, implementation, and maintenance (Corcho et al., 2005). Support activities include knowledge acquisition, integration, evaluation, documentation, and configuration management throughout the entire process of developing ontologies (Corcho et al., 2005). The three groups of activities are divided into smaller activities and each activity is further divided into small tasks.

The five stages of the development activities (i.e., specification, conceptualization, formalization, implementation, and maintenance) are the focus of Study 2 and illustrated in Figure 4.4.

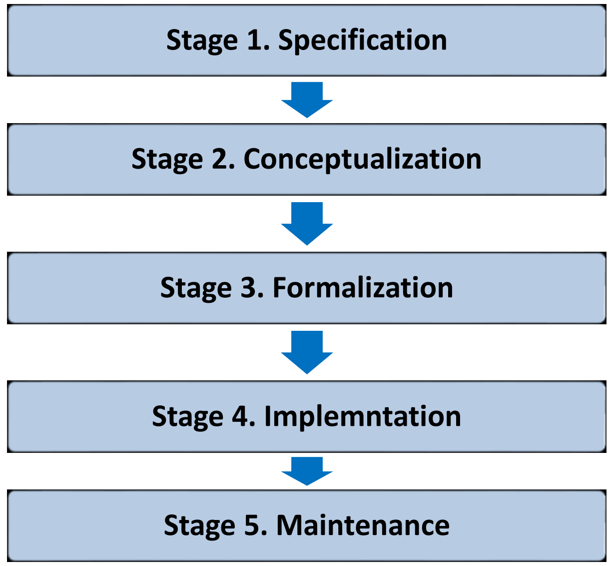


Figure 4.4 Ontology building stages of methontology

Since knowledge acquisition mainly occurs in the specification and conceptualization stages in the development activities, knowledge acquisition is important to Study 2 in collecting domain knowledge for building ontologies. Each stage of the development activities is especially associated with knowledge acquisition, as will be explained in detail.

4.2.1. Stage 1. specification

Specification involves creating a collection of "requirements that the ontology should fulfill" (Suárez-Figueroa, Gómez-Pérez, & Villazón-Terrazas, 2009, p. 982). In methontology, ontology requirement specification covers three major aspects: the purpose of the ontology, intended users, and ontology requirements that the ontology should satisfy after being formally implemented.

To identify ontology requirements, various methods can be used, including brainstorming, joint application development, exploit scenario, use cases, interviews with users and domain experts, and competency questions. Among these methods, competency questions are selected because they are natural language questions that express "a pattern for a type of question people expect an ontology to answer" (Ren et al., 2004, p. 752). Competency questions help capture scope, content, and other conditions. For example, competency questions for family history may include questions that must 1) represent marriage; 2) represent kin relations or other questions. Competency questions were first proposed for building an ontology by Grüninger and Fox (1995) and adopted by several ontology building methods, such as On-to-Knowledge (Sure, Staab, & Studer, 2004), the Unified methodology (Uschold, 1996), and Ontology Development 101 (Noy & McGuinness, 2001). Thus, Study 2 adopts competency questions to identify ontology requirements. In addition, Study 2 uses the entities and categories that are identified in Study 1 and existing archival glossaries.

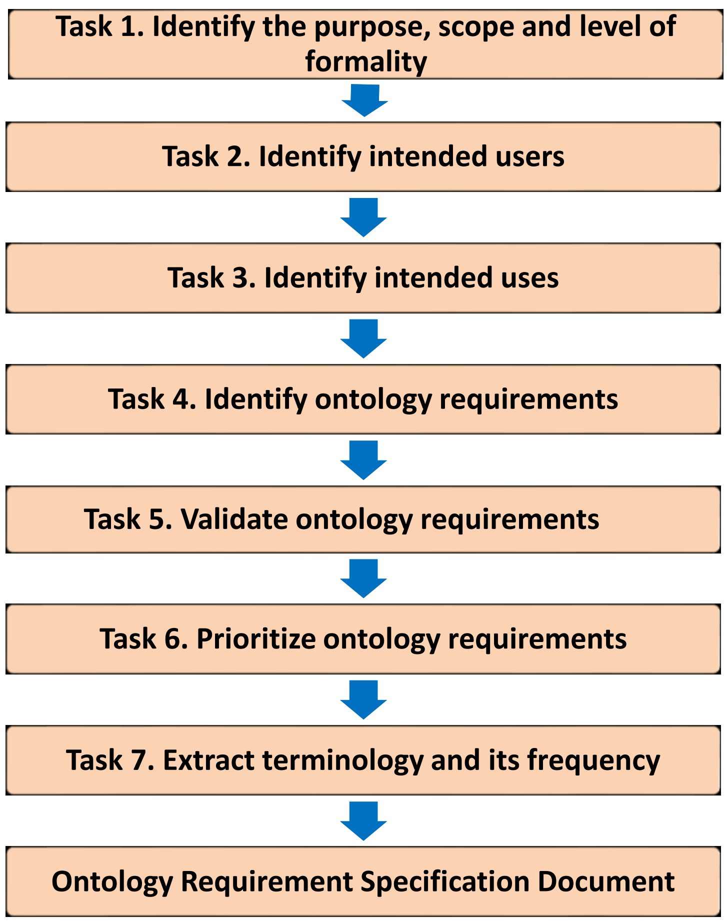


Figure 4.5 Tasks in the specification stage

The specification stage aims to identify the definition, goal, inputs, outputs as well as when and who (i.e., the researcher) carried out what activity at this stage. To do so, the activities of this stage are divided into seven tasks, as illustrated in Figure 4.5.

Task 1. Identify the purpose, scope, and level of formality.

This task is to determine the main purpose of the ontology, its scope, the level of granularity and the degree of formality that ranges from informal natural language to a rigorous formal language.

Output: purpose, scope, and the level of formality;

Task 2. Identify intended users.

This task is to determine intended users that will use the ontology.

Output: a list of intended users

Task 3. Identify intended uses.

This task is to determine what kinds of scenarios the ontology will be used for. Since UML (Unified Modeling Language) is a standard language for developing software blueprints (Rumbaugh et al., 1998), it has the advantage of visualizing, specifying and constructing models that can be understood by non-expert users. Thus, a use case diagram of UML was utilized to capture its intended uses.

Output: a list of intended uses in the form of scenarios; UML-based (Unified Modeling Language) four use cases (see APPENDIX XI. Use cases)

Task 4. Identify the ontology requirements of archival descriptions.

This task is to collect requirements and write the requirements in natural language in the form of competency questions. In order to identify important concepts from answers to competency questions, a mind map tool is used to visually organize concepts. A mind map is hierarchical and shows relationships among the pieces of the whole. It makes it easier to manage questions. The questions were developed based on the four use cases developed in Task 3. From the use cases, questions are categorized into four groups that cover archival context, archival records, the way in which archivists process records, and creators of records. Each group has questions about a specific aspect of archives.

Output: a list of competency questions (see Table 4.6 section 6)

Task 5. Validate ontology requirements.

This task is to identify the potential problems (e.g., conflicts and overlooked questions) of the listed competency questions. The following criteria are adopted by: 1) correctness; 2) completeness; 3) consistency; 4) verifiable; 5) understandable; 6) no ambiguity; 7) conciseness; 8) realism; 9) modifiable; 9) traceable (Davis, 1993).

Output: a confirmation about the validity of the competency questions

Task 6. Prioritize ontology requirements.

This task is to set different priority levels for each competency question.

Output: a list of priorities attached to each competency question

Task 7. Extract terminology and its frequency.

This task is to extract terminology and its frequency. A pre-glossary of terms has been developed (see APPENDIX XII. Terminology and frequency).

Output: a pre-glossary of terms

As a result, the following ontology requirement specifications are developed in Table 4.6.

	Archival Description Ontology Requirement Specifications
1	Purpose
	Provide a consensual knowledge model of archival context that could be used by archivists, researchers, and other users.
2	Scope
	The ontology focuses on archival context (e.g., entities, events, and relations). The level of granularity is directly related to the competency questions and identified terms.
3	Level of Formality
	The ontology has to be implemented in OWL.
4	Intended Users
	User group 1. Archivists who create and manage archival context. User group 2. Researchers who search and use archival context. User group 3. Normal users who search and use archival context occasionally.
5	Intended Uses
	Use case 1. Create archival descriptions (See APPENDIX XI. Use cases - Use case 1) Use case 2. Manage archival descriptions (See APPENDIX XI. Use cases - Use case 2) Use case 3. Browse archival descriptions (See APPENDIX XI. Use cases - Use case 3) Use case 4. Search archival descriptions (See APPENDIX XI. Use cases - Use case 4)
6	Competency Questions
	Group 1. Questions about archival context 1) What entities are components of archival context? 2) What entities are used to provide for information on archival context? 3) What are the features of these entities? 4) Can they be grouped into identifiable categories? 5) What are the relations of those entities in representing archival context? 6) Are there other essential categories? 7) What is the archival context of this archival description? 8) Who produced this archival context? When? At which location? Group 2. Questions about archival records What components are considered the archival context of files, folders, objects, series, and fonds by archivists, researchers, and regular users? Group 3. Questions about how archivists process archives How is the archival context of records created, managed, and used? Group 4. Questions about records' creators What components are included in the archival context related to creators?
7	Pre-Glossary of Terms
	Terms Frequency (See APPENDIX XII. Terminology and frequency)

Table 4.6 Ontology requirement specifications

As identified in this specification stage, the answers to the competency questions were organized into a pre-glossary of terms. The answer to each question is analyzed, and the main terms were compiled into an Excel spreadsheet. The frequency of each concept was calculated. The terms and frequencies were listed as a pre-glossary of terms.

4.2.2. Stage 2. conceptualization

Conceptualization is organizing and structuring knowledge acquired from external sources. Conceptualization activities are divided into a set of smaller tasks. Each task converts informal data into semi-formal representations that are easily understood by both domain experts and ontology developers.

Conceptualization and knowledge acquisition are closely connected since knowledge acquisition largely occurs in the specification and conceptualization stages as explained above. In this study, knowledge acquisition has three major sources, including 1) entities, relations, and events that are identified in the content analysis of Study 1; 2) the pre-glossary of terms that are generated from the specification stage; 3) a compiled glossary of terms that are collected from authorized standards (Bureau of Canadian Archivists, 2008; International Council on Archives, 2010; The InterPARSE project team, 2001). There are overlaps among the three sources, each of the three glossaries presents a perceived view of archives, and the combination of the three sources can minimize the risk of missing out on any important aspect of the domain.

At this stage, concepts, relations, instances, and attributes are first identified. These form the base knowledge to develop formal axioms and rules. Concepts are taken in a broad sense in intermediate representations. For instance, concepts of archives (e.g., person, organization, and fonds) were identified. Relations represent a type of association between the concepts of the domain (Corcho et al., 2005). If the relation links two concepts, it is called a binary relation. For example, membership is a relation between a person and an organization. Instances represent a specific person, organization, or fonds. For example, Bill Gates is an instance of a concept *Person*. Relations can be instantiated between two instances. In the following example, the *isStepsonOf* relation is instantiated between two individuals: Harold J. Smith (*Person*) and James L. G. Fitz Patrick (*Person*).

> [Example] The vast majority of the collection was donated to the College of Staten Island by James L. G. Fitz Patrick's stepson, Harold J. Smith, in 2000.
>
> Relation: *isStepsonOf* (Harold J. Smith, James L. G. Fitz Patrick)

Attributes describe properties of instances and concepts. Instance attributes describe concept instances (e.g., the first name of a person). Class attributes describe concepts (e.g., type of control is a class attribute of the concept company. Private, public, shared controls are three types of the company). Formal axioms are logical expressions that are always true and normally used to specify constraints in the ontology. Rules are used to infer knowledge in the ontology, such as attribute values and relation instances.

The activities of the conceptualization stage are divided into eleven tasks, as illustrated in Figure 4.6.

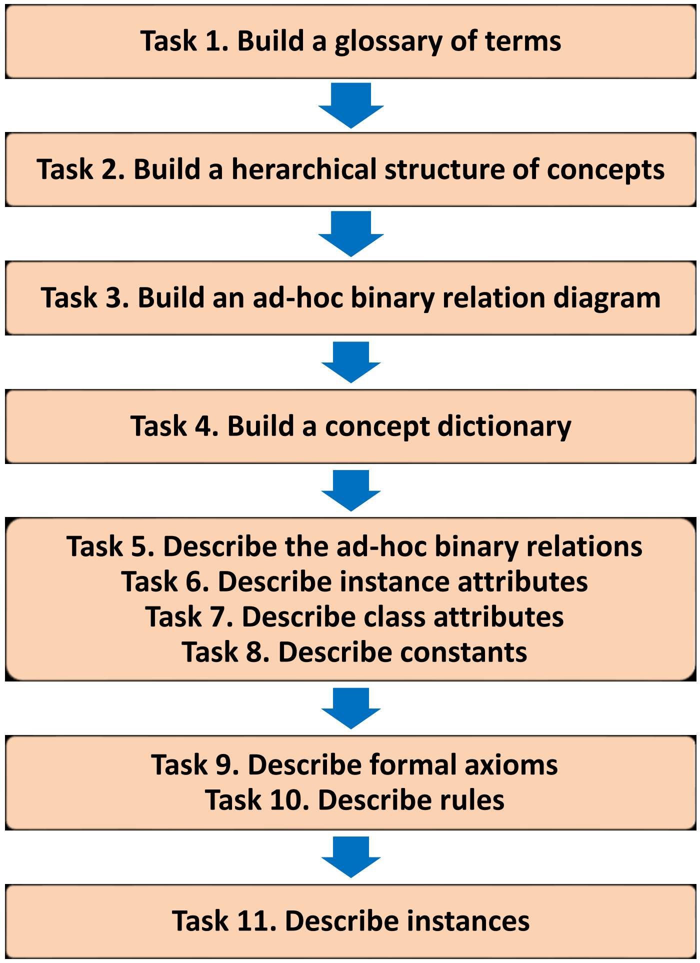


Figure 4.6 Tasks in the conceptualization stage

The tasks should be performed in a sequence. In particular, the output of Task 1 is an important input for the rest of the tasks. If new vocabulary is introduced, any previous tasks need to be revisited. Each task of the conceptualization stage is explained in detail.

Task 1: Build a glossary of terms.

A glossary of terms includes all the relevant terms from three sources (i.e., results of Study 1, pre-glossary of terms created in the specification stage, and a compiled glossary based on authorized sources (Bureau of Canadian Archivists, 2008; International Council on Archives, 2010; The InterPARSE project team, 2001). In addition to the results of Study 1, in order to broaden sources of knowledge acquisition and mitigate possible biases in the knowledge acquisition, existing sources were added, including the terminology database developed by the InterPARES project (The InterPARES project team, 2001), the glossaries found in Rules for Archival Description (RAD) (Bureau of Canadian Archivists, 2008, Appendices D) and the International Standard on Archival Description (General) (ISAD(G)) (International Council on Archives, 2010, p. 10) as the primary sources of the methontology study. Terms from InterPARES, RAD, and ISAD(G) are compiled into an Excel spreadsheet. Then terms are sorted, and similar terms are merged. The frequency of each term is calculated. Any term with a frequency score of 2 or more is compiled into a list of terms. Similarly, all of the pre-glossary of terms and terms from results of Study 1 are put into an Excel spreadsheet. The frequency of each term is calculated. Any term with a frequency score of 2 or more is kept in the glossary of terms. Each entry in the glossary consists of name, natural language description, and type. Type is selected from concept, attribute, and relation. An excerpt of the glossary of terms is shown in Table 4.7 (See the full list in APPENDIX XIII. A glossary of terms).

Term	Description	Type
Access	The right, opportunity, or means of finding, using or approaching documents or information.	Concept (Right/Event)
Accession	The process of formally accepting and recording the receipt of records into archival custody.	Concept (Event)
...	…	…
Brother	A man or boy in relation to other sons and daughters of his parents	Relation (Family Relation)
Building	A property	Concept (Property)
Bury	Event	Concept (Event)
Business	Company	Concept (Agent)
Capture	Event	Concept (Event)
Cause	Causal Relation	Relation
City	A place	Concept (Place)
…	…	…

Table 4.7 An excerpt from the glossary of terms

Each identified concept is assigned to a proposed category of the type into which the concept falls. Some terms have multiple meanings so there may be more than one proposed category. At this stage, although some terms in the glossary of terms may be too general or specific and refer to the same term, the aim is to include relevant terms whenever possible and the glossary of terms is a starting point to further define concepts and build a formal, hierarchical structure of concepts.

Task 2: Build a hierarchical structure of concepts.

A glossary of terms is the primary source for building the hierarchical structure of concepts. In methontology, to construct a hierarchical structure, three approaches are used: top-down, bottom-up, and middle-out approaches. A top-down approach is to start from the high-level concepts and then find the lower levels of concepts, while a bottom-up approach starts from the bottom classes to the super-classes. A middle-out approach is a combination of top-down and bottom-up approaches. This study took the middle-out approach. In the top-down approach, the ten classes of contextual entities that are identified in Study 1 are taken as key concepts: *Agent, Artifact, Event, Feature, Function, Place, Relation, Role, Situation*, and *Time*. Among these, *Relation* is a special entity and can be considered as an object property of an ontology. Similarly, *Feature* is also a special entity and can be considered as a data property of an ontology. Since both object and data properties are integral parts of an ontology, they are not included as the top key entities. Therefore, the rest of eight entities are the contextual entities (i.e. *ContexutalEntity* class in Figure 4.7).

In the meantime, a bottom-up approach is to cluster and group concepts. For example, in the glossary of terms, terms like *Author*, *Creator*, and *Donor* indicate different roles that individuals, families or corporate bodies may have. Thus, these terms can be grouped under the concept *Role*. At the same time, the concept *Role* can be placed under the concept *Abstraction*. Similarly, *Occurrence* is the super-class of the *Event* and *Situation* classes. After this process, five direct sub-classes of the class *ContextualEntity* are selected, as shown in Figure 4.7. The sub-class hierarchical structure is also displayed in Figure 4.7 as well.

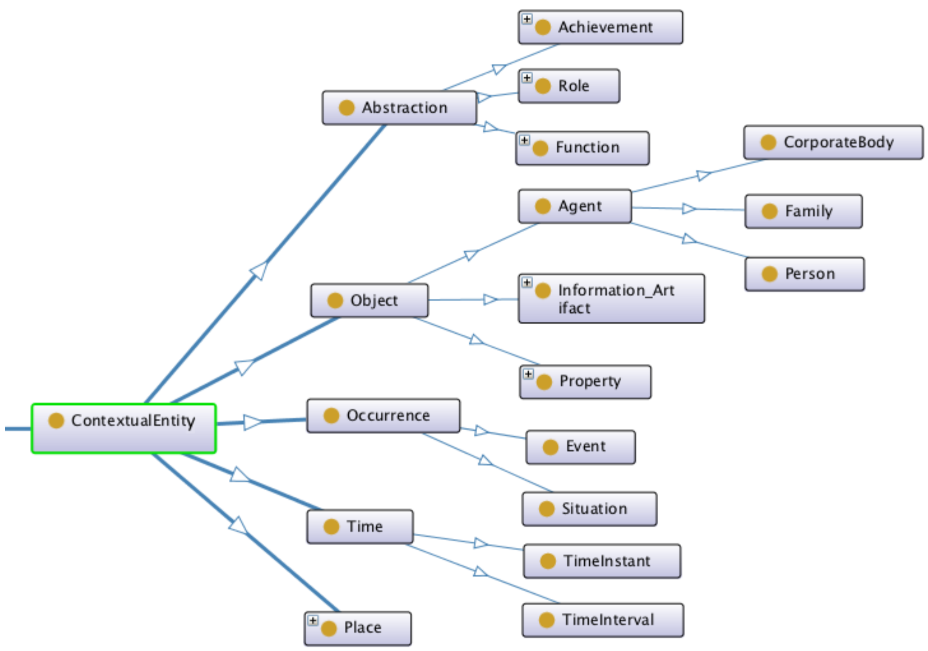


Figure 4.7 Concept structure with sub-classes

The concept of *ContextualEntity* has five direct sub-classes: *Time*, *Place, Occurrence*, *Object,* and *Abstraction*. For example, as illustrated in Figure 4.7, the entity *Agent* is a type of *Object*, which is placed under the entity of *Object*. The entity *Role* is a type of *Abstraction*, which is a lower level of the entity. The concepts (called as classes) listed in Figure 4.7 are derived from the glossary of terms (see APPENDIX XIII. A glossary of terms). For instance, *Function* is a concept directly from the glossary. *Time* is the concept that is derived from the date related terms. *Place* is the concept that is from the terms: *City*, *County*, *Region*, *State*, and *County*. Classes, such as *Situation* and *Event* under *Occurrence*, *Role*, *Achievement* and *Function* under *Abstraction,* were converted from concepts directly from the concepts dictionary.

In Study 2, for the hierarchical structure, four relations are used: subclass-of, disjoint-decomposition, exhaustive-decomposition, and partition (Horridge et al., 2011). A subclass-of relation is to indicate that a concept *X* is a subclass-of another concept *Y* and only if every instance of *X* is also an instance of *Y*. For example, a concept *Person* is a subclass-of another concept *Agent*. A disjoint-decomposition of a concept *X* is a set of subclasses of *X* that does not have common instances and does not cover *X*. For example, concepts *Person* and *Corporate Bod*y do not have common instances because an instance of a *Person* cannot be an instance of *Corporate Body* and they are two subclasses of *Agent*. Meanwhile, there are other instances (e.g., instances of *Family*) that are neither instances of *Person* nor *Corporate Body*. The concepts *Person* and *Corporate Body* are a disjoint-decomposition of the concept of *Agent*.

An Exhaustive-Decomposition of a concept *X* is a set of subclasses of *X* that cover *X* and may have common instances and subclasses. For example, an exhaustive decomposition of the concept *Person* is a set of three concepts, including *Adult*, *Student*, and *Preschool Child*.

A partition of the concept *X* is a set of a subclass of *X* that does not share common instances and covers *X*, which means that they are not instances of *X* and that are not instances of one of the concepts in the partition. For example, a partition of the concept *Person* is a set of two concepts, including both concepts *Adult* and *Child*, because there is no person that can be an *Adult* and *Child* at the same time. The concepts *Adult* and *Child* cover the concept *Person* as all instances of *Person* are instances of either *Adult* or *Child*. Between classes that are used in Figure 4.7, there are more specific relations (i.e., disjoint-decomposition, exhaustive-decomposition, and partition). The classes *Time*, *Place, Occurrence*, *Object*, and *Abstraction* are disjoint with each other but the disjoint relations are not explicitly presented in the concept structure. The four types of relations need to be examined to avoid errors. For example, an instance cannot be simultaneously an instance of classes of a disjoint-decomposition. Loops should not be in the hierarchical structure, so it is critical to ensure that there is no loop in the structure.

Task 3: Build an ad-hoc binary relation diagram.

The ad-hoc binary relations are to further connect concepts that are included in the hierarchical structure. For example, the concept *Agent* is not directly linked to the concept *Role* in the hierarchical structure because there is no sub-classs/super-class relation between the two concepts. However, the relation *hasRole* and its inverse relation *isRoleOf* can connect the two concepts (i.e., *Agent* and *Role*). The concept *Agent* is the domain, and the concept *Role* is the range of the relation *hasRole.* Similarly, the concept *Role* is a domain, and the concept *Agent* is the range of the relation *isRoleOf.*

Errors in ad-hoc binary relations are in the domains (a domain can be considered as a source concept) and ranges (a range can be considered as a target concept). Domains and ranges need to be precise but not overly specific. Table 4.8 displays an excerpt of the ad-hoc binary relation table with the name of the relation, domain, and range. The full ad-hoc binary relation table is included in APPENDIX XIV. A complete ad-hoc binary relation table.

Name of Relation	Domain	Range
associate		
before_evt	Event	Event
has Function		Function
has lived at	Person	Location
hasFeature	Entity	

hasMilitaryStatus	Person	MilitaryStatus
hasOccupation	Person	Position
hasRole	Agent	Role
hasSource		Publication
hasStatus		Status
hasSubject		Subject
LocatedAt		Location

Table 4.8 An excerpt of ad-hoc binary relations

For example, the *hasOccupation* relation connects a *Person* with a *Position*. However, some relations have no specific domain or range, as shown in Table 4.11. The domain and range of the associated relations are not specified, which means that the relation has no specific domain or range. In other words, the associated relation can exist between any two entities (e.g., *Person* and *Person*, *Event* and *Person*, etc.).

Task 4: Build a concept dictionary.

The concept dictionary consists of all concepts, their relations, instances, and attributes. In methontology, instances are optional. In building the concept dictionary, the concepts in the hierarchical structure and their relations from the relation table are combined into one concept dictionary. Concepts are checked with their relations to see whether the concepts are either the domain or range of the relations as a validating process. An excerpt of the concept dictionary (the full dictionary see APPENDIX XV. A concept dictionary) is shown in Table 4.9.

Concept Name	Instances Attributes	Relation
Time		occurredAtTime
Location		occurredAtPlace partOf associate

Event		occurredAtTime occurredAtPlace precedeSituation followSituation before_evt during_evt start_evt finish_evt meet_evt overlap_evt
Agent		associate creatorOf hasAttributehasAttribute
Person	Name Birthday Place of Birth Day of Death	isFatherOf isMotherOf hasFather hasMother hasBrother hasSister
Abstract		hasAttribute associate
Function		hasFunction hasAttribute associate
Role		hasRole hasAttribute associate

Table 4.9 An excerpt of the concept dictionary

Task 5: Describe ad-hoc binary relations.

To explain the details of each relation, the cardinality constraint and its inverse relation need to be included for each ad-hoc binary relation. A cardinality constraint expresses restrictions on the overall number of values of a relation (Schreiber, 2005). For example, a *hasParent* relation can have a cardinality constraint to indicate that a person can have exactly two persons as his/her parent. The following is a portion of an ontology:

Class: Person
Annotations: rdfs:label "Person"@en

SubClassOf: hasAge exactly 1
and hasGender exactly 1
and hasGender only {female, male}
and hasParent exactly 2

In the example, it defines a class *Person,* which has exactly one *hasAge* and hasGender property (relation) respectively and two *hasParent* properties. Every relation has a direction, from the domain to the range. For example, an archive's (source concept) own records (target concept). An inverse relation of a relation defines the inverse direction of the relation. The previous sentence can be expressed as: records are owned by archives. Thus, the ad-hoc binary relation table is further expanded to the full ad-hoc binary relation table, as displayed in Table 4.10 (an excerpt from APPENDIX XIV. A complete ad-hoc relation table).

Name of Relation	Domain (Source Concept)	Source cardinality	Range **(Target Concept)**	Inverse Relation
HasFunction			Function	isFunctionOf
Has lived at	Person		Location	
hasFeature	Entity			isFeatureOf
hasMilitaryStatus	Person		MilitaryStatus	isMilitaryStatusOf
hasOccupation	Person		Position	isOccupationOf
hasRole	Agent		Role	isRoleOf
hasSource			Publication	isSourceOf
hasStatus			Status	isStatusOf
hasSubject			Subject	isSubjectOf
LocatedAt			Location	
Associate				Associate
Before_evt	Event		Event	After_evt

Table 4.10 An excerpt of the full ad-hoc binary relation table

Task 6: Describe instance attributes.

This task is to describe instance attributes included in the concept dictionary. Instance attributes are attributes of the instances of concepts. Their values may be different for each instance of the concept. Each entry of the instance attributes table corresponds to one instance attribute. For each entry, there is a list of fields: the name of an instance attribute, the concept it belongs to, value type (e.g., integer, string, date, or boolean), and value range (if applicable), minimum and maximum cardinality (e.g., a statement "a person has one biological father" indicates that the minimum and maximum cardinalities of the attribute (has biological father) are exactly one), other elements if applicable (e.g., class attributes, instance attributes, and constants) used to infer values of the attribute. Table 4.11 shows one example of instance attributes for the concept of *Person* by gender and name.

Name of an instance attribute	Concept	Value Type	Value Range	Cardinality (min, max)
gender	Person	An enumerated type can be either male or female in this example.	male, female	(1, 1)
surname	Person	String	-	(1, 1)

Table 4.11 An example of the instance attribute table

Task 7: Describe class attributes.

This task describes all of the class attributes included in the concept dictionary. Class attributes provide additional information about a concept. Class attributes are neither inherited by the subclasses, nor by the instances. Each entry of the class attribute table consists of a detailed description of the class attribute that includes a list of fields: the name of the class attribute, the concept it belongs to, value type, value(s), cardinality, and the instance attributes whose values can be inferred with the value of the class attribute. Table 4.12 shows an example of the class attribute *age* that is defined for the concept *person.*

Name of a class attribute	Concept	Value Type	Value	Cardinality (min, max)
Age	Person	Integer	0-200	(1, 1)
Year of Birth	Person	Integer		(1, 1)

Table 4.12 An example of the class instance attribute table

Task 8: To describe constants.

This task describes each of the constants defined in the glossary of terms. Each entry of the constant table should consist of a detailed description of a constant, including the name of the constant, value type, value, and the attribute that can be inferred using the constant. Table 4.13 is an example of a constant.

Name	Value Type	Value	Attribute
The age of majority in Ontario	Number	18	-

Table 4.13 An example of the constant table

Task 9: Describe formal axioms.

This task identifies the formal axioms that are needed for the ontology and describes them precisely. Once concepts and their hierarchical structure, ad-hoc relations, attributes, and constants have been defined, methontology proposes describing formal axioms and rules in parallel. However, as this study focuses on developing a semi-formal archival context ontology, it does not fully describe formal axioms and rules at this stage.

Task 10: Describe rules.

This task identifies the rules that are needed for the ontology and then describes them. In methontology, rule expressions use *if* <*conditions*> *then* <*consequence*>. The <*conditions*> consist of conjunctions of atoms, and the <*consequence*> is a single atomic condition.

Task 11: Describe instances.

Each instance should be defined with a list of fields, including a name, the concept it belongs to, and its attributes. Table 4.14 presents some instances of essays, reports, and others, which are instances of the concept *information artifact*. These instances could have different attributes. For example, correspondence is a type of information artifact mainly created between two persons.

Instance name	Concept	Attribute
Essay	Information_artifact	has Author
Report	Information_artifact	relatedtoPerson
Memorandum	Information_artifact	relatedtoPerson
Minute	Information_artifact	relatedtoMeeting
Correspondence	Information_artifact	hasAuthor hasReceiver

Table 4.14 An example of the instances table

Going through the eleven tasks in the procedure, the outcome of conceptualization is to build a conceptual model enough to represent archival contexts. Among the entities in the concept hierarchical structure, Lee (2011) remarked that "the "*Occurrence*" is a "characterization, for a given span of times and places, of either the state of a set of entities or their interaction(s)" (p. 106). The conceptual model behind the various tables can embody the occurrences. They are carried out by agents using other objects that play various roles or fulfill functions for a given span of times and places. Thus, occurrences are the core of this archival context conceptual model. Its initial conceptual structure is shown in Figure 4.7.

4.2.3. Stage 3. formalization

Formalization is formalizing the results of conceptualization and turning the results into a formal knowledge representation language (Gómez-Pérez, et al., 2004). Through the graphical user interface of Protégé (an ontology editing tool), the conceptual model of archival context can be formalized and constructed into an ontology. During the construction of an ontology, there is a process to double-check sub/super types, relations, and definitions of classes and relations. Protégé supports imports and exports to different formats, such as OWL/XML and RDF/XML. This study built an ontology by using OWL/XML language. Table 4.15 displays an excerpt of it and the complete version is listed in APPENDIX XVI. A complete event ontology for archival context.

```xml
<?xml version="1.0"?>
<Ontology xmlns="http://www.w3.org/2002/07/owl#"
 xml:base="http://archdesc.info/eoac"
 …
 ontologyIRI="http://archdesc.info/eoac">
 <Prefix name="" IRI="http://archdesc.info/eoac#"/>
 …
 <Annotation>
      <Literal datatypeIRI="http://www.w3.org/2001/XMLSchema#string">A core ontology created by Jason Zou
      for describing archival description.</Literal>
 </Annotation>
 <Annotation>
      <AnnotationProperty abbreviatedIRI="rdfs:label"/>
      <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Event Ontology for
      Archival Context (EOAC).</Literal>
 </Annotation>
 <Declaration>
       <ObjectProperty IRI="#associatedWith"/>
 </Declaration>
 …
</Ontology>
```

Table 4.15 An excerpt of the ontology for archival contexts in OWL and RDF

Based on the outcomes that have been identified in the previous procedure, including concept dictionary, hierarchical structure, relations, and conceptual model, an archival context ontology was built with five top classes: *Time*, *Place*, *Occurrence*, *Object*, and *Abstraction*. The root class in OWL is a thing (owl:Thing) which is the root of all classes. The five classes are sub-classes of the *ContextulEntity* class, which is a sub-class of owl:Thing. In the formalization/implementation, the interrelations and characteristics of classes and relations are further determined and defined using the OWL property and ontological restriction, which are presented in the next sections.

4.2.3.1. **Time**

The class Time has two sub-classes: *TimeInterval* and *TimeInstant* to present the temporal dimension of occurrences. The *TimeInterval* is a class to model things with extent. The *TimeInstant* class is to model “point-like in that they have no interior points” type of time. For example, there was a meeting at 6 pm on May 1, 2010. “6 pm on May 1, 2010” is an instant. An instant can be considered "an interval with zero length, where the beginning and end are the same" (Hobbs & Pan, 2006, sec. Topological Temporal Relations). For example, an interval can be a couple of seconds, minutes, or even decades.

4.2.3.2. **Place**

Place is a concept to present the geospatial dimension of an ontology. A Place can be a county, township, city, region, or country. The initial version of the place class is directly derived from the concept dictionary and shown in Figure 4.8.

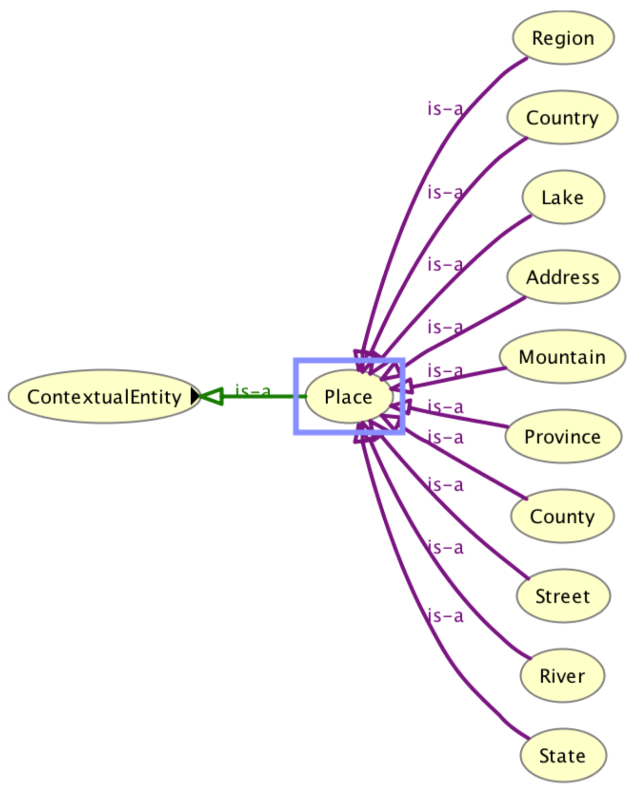


Figure 4.8 The class place

A place with different names is common. The name of a place may be changed due to various reasons. The changes present additional information on the external context. To accommodate those scenarios, the initial version is not adequate.

4.2.3.3. **Occurrence**

Occurrences refer to “entities that occur or happen” (Arp et al., 2015, p.87). Based on the situation calculus (McCarthy & Hayes, 1969), in this study, occurrence has two main classes: *Situation* and *Event*. A situation provides the “context for framing time-dependent properties of (possibly multiple) entities” (Logoze & Hunter, 2001, p. 162). An event is an entity that describes certain changes or experiences. An event marks a “transition from one situation to another” (Logoze & Hunter, 2001, p. 162). For example, an establishment event shows the relation between one event and two situations.

In 1889, Acadia University … **[formed]** the Acadia Amateur Athletic Association.

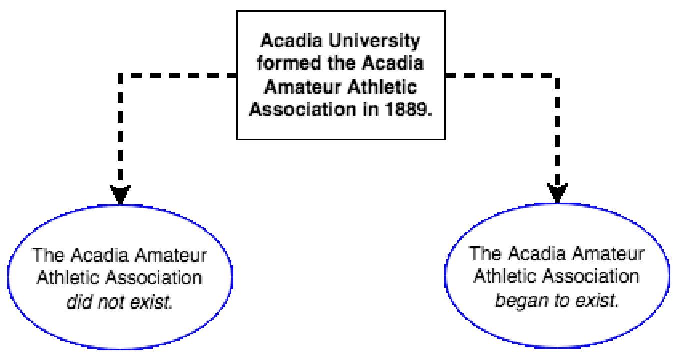


Figure 4.9 An event and two situations

In Figure 4.9, the establishment *event* denoted in the rectangle implies two situations in two blue ovals: 1) The Acadia Amateur Athletic Association did not exist; 2) The Acadia Amateur Athletic Association began to exist. The two relations (i.e., *precede* and *follow*) are used to connect the event and the two situations. For instance, the above example can be modeled as:

<The Pre-Situation> *precede* <The Establishment Event>
<The Post-Situation> *follow* <The Establishment Event>

If this is done manually, it brings increases the likelihood of errors by those who annotate.

In the previous specification and conceptualization stages, the concept dictionary and relation tables have been created, and different types of events have been identified. Events can be grouped by theme (e.g., *LifeEvent*, *OrganizationalEvent*, *ArchivalEvent*, etc.). At the same time, events can also be grouped by types of changes (e.g., *temporalChangeEvent*, *spatialChangeEvent*, *cardinalChangeEvent*, and *identityChangeEvent*). Theme and change-oriented events can co-exist. Figure 4.10 shows the initial version of the event class.

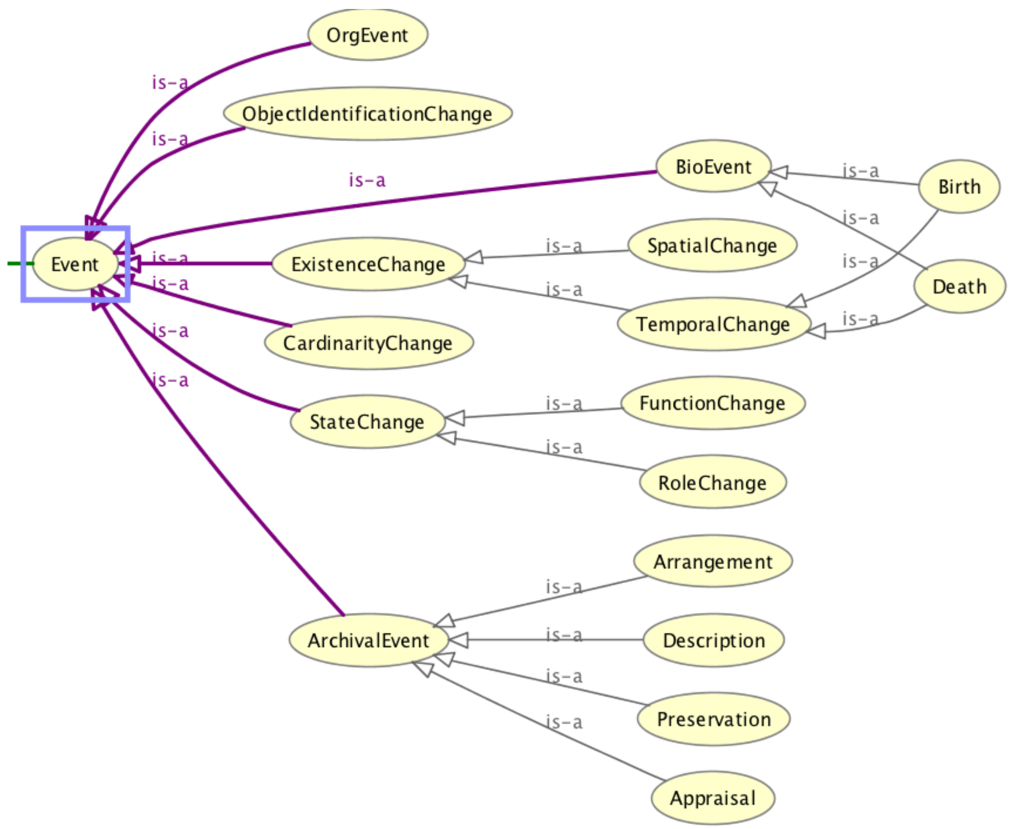


Figure 4.10 The class *event* (partial view)

In Figure 4.10, the Birth event is not only a *BioEvent* but also an *ExistenceChange* Event. A *TemporalChange* event refers to the existence of an object in such event changes when there is a change in time. For a *Birth* event, when there is a change in time, a person is born (from not existing to existing outside of his/her mother).

4.2.3.4. **Object**

Object refers to concretely tangible entities, such as a computer, a car, and a table. This object class has three main sub-classes: *Information_Artifact*, *Agent* and *Property*. Figure 4.11 shows a partial view of the class.

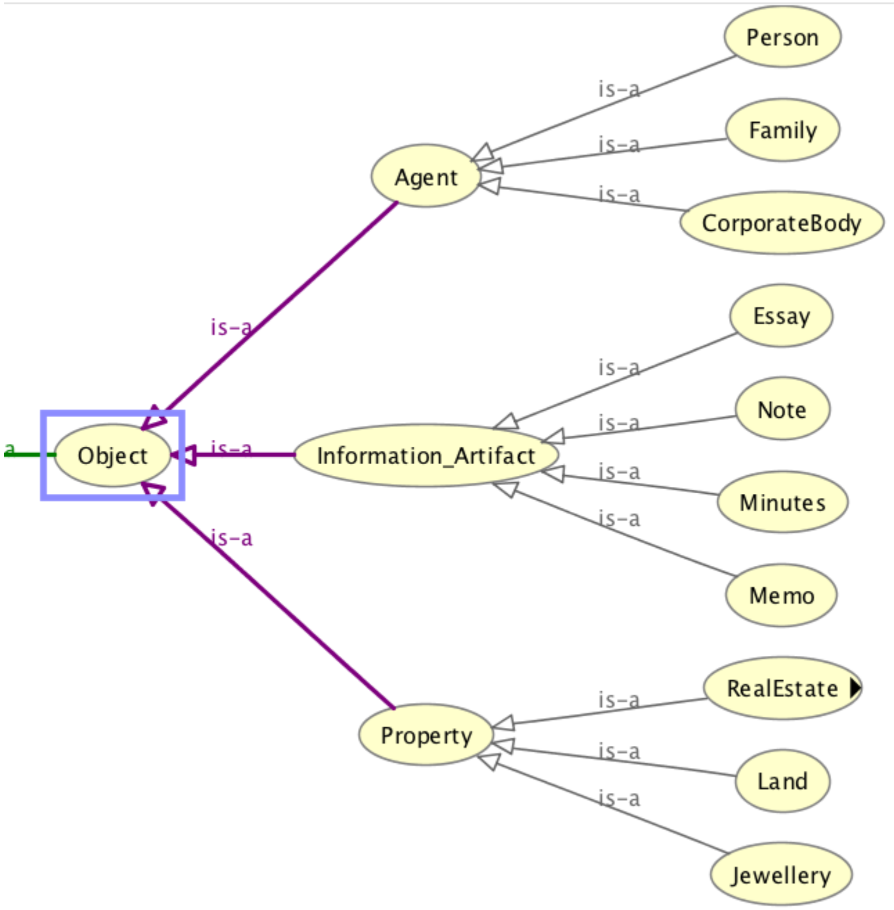


Figure 4.11 The class *object* (partial view)

4.2.3.5. **Abstraction**

Abstraction is a class to model entities that are not concrete things, such as achievements, roles, and functions. Figure 4.12 displays the abstraction class.

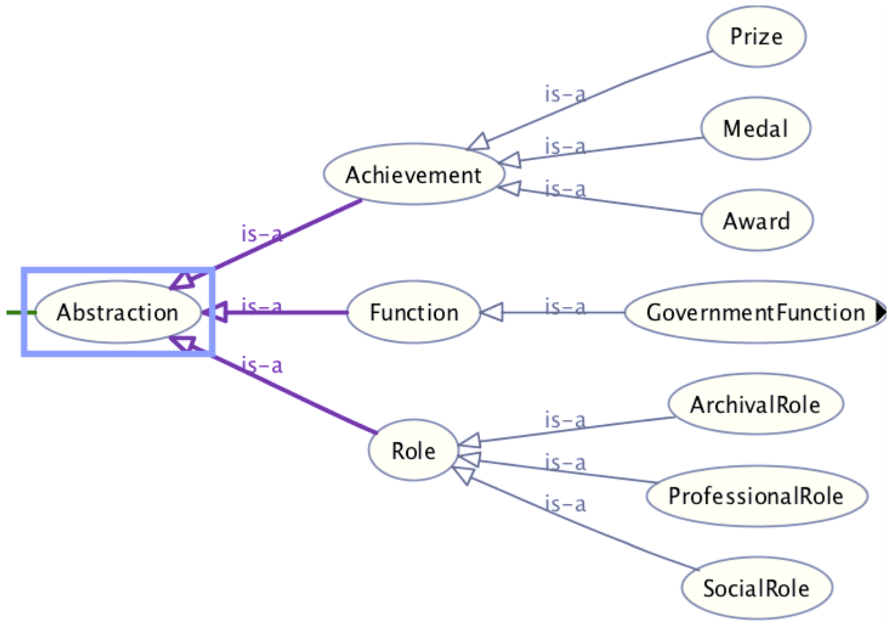


Figure 4.12 The class *abstraction* (partial view)

4.2.4. Stage 4. implementation

Through the previous procedure, the concepts, their relations and attributes, and the conceptual model have been identified. They have also been presented in natural language or intermediate format. At the implementation stage, all should be converted into formal knowledge representations. Implementation is to "formally represent and implement the products" of a specification and conceptualization by using an ontology development environment (Gašević et al., 2006, p. 67). In this stage, concepts, relations, and attributes are expressed in an ontology language, OWL. The structure of the concepts is organized in super-classes and sub-classes of the hierarchy.

Protégé is a "free, open source ontology editor and framework for building intelligent systems" (Stanford Center for Biomedical Informatics Research, 2016) and has been widely used in building ontologies. It is used as a primary tool for implementation in this study. This implementation stage is divided into the five detailed tasks shown in Figure 4.13 as follows:

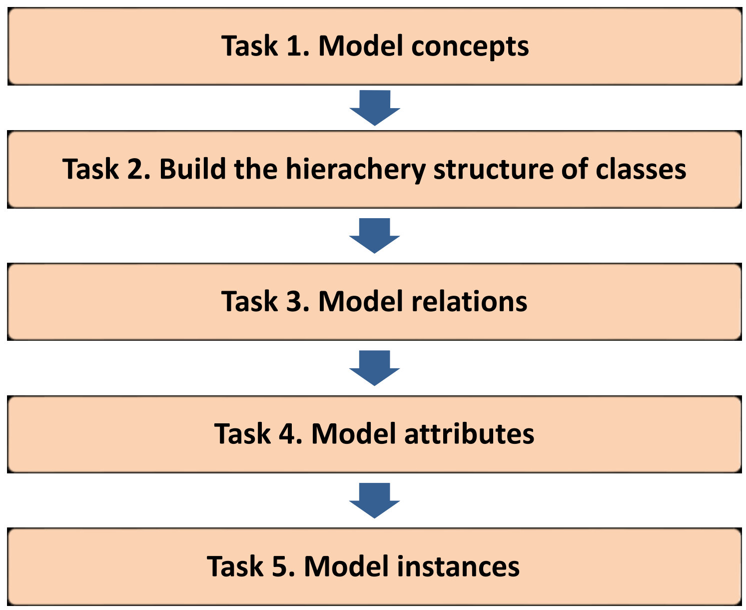


Figure 4.13 Tasks in the implementation stage

Task 1: Model concepts.

Since classes in an ontology are also called concepts, concepts in the concepts dictionary can be directly converted to classes in Protégé. In this task, classes were created and defined, based on the concepts dictionary.

In Protégé, Task 1 can be done by using the “Classes” tab, as illustrated in Figure 4.14, which is for modeling classes. The “Classes” tab is divided into three areas (marked by yellow circles). In area 1, the "Class hierarchy" displays all classes in a tree-type structure. Classes can also be added, modified, and deleted in Area 1, where the position of a class in the hierarchy can be edited as well. In area 2, annotations (notes) of the concept can be added, modified and deleted. There are nine default annotation properties (i.e., owl:backwardCompatibleWith, owl:deprecated, owl:incompatibleWith, owl:versionInfo, owl:priorVersion, rdfs:comment, rdfs:isDefinedBy, rdfs:label, and rdfs:seeAlso) that can be used to define the class and provide version control, compatibility, and note information on the class. The second tab, “usage,” lists all of the usages of the concept. The areas 1 and 2 are sufficient for this task. Area 3 is used in Task 2 for modeling hierarchical structure and Task 5 for creating instances. Figure 4.14 shows the class *Event* in Protégé, as follows:

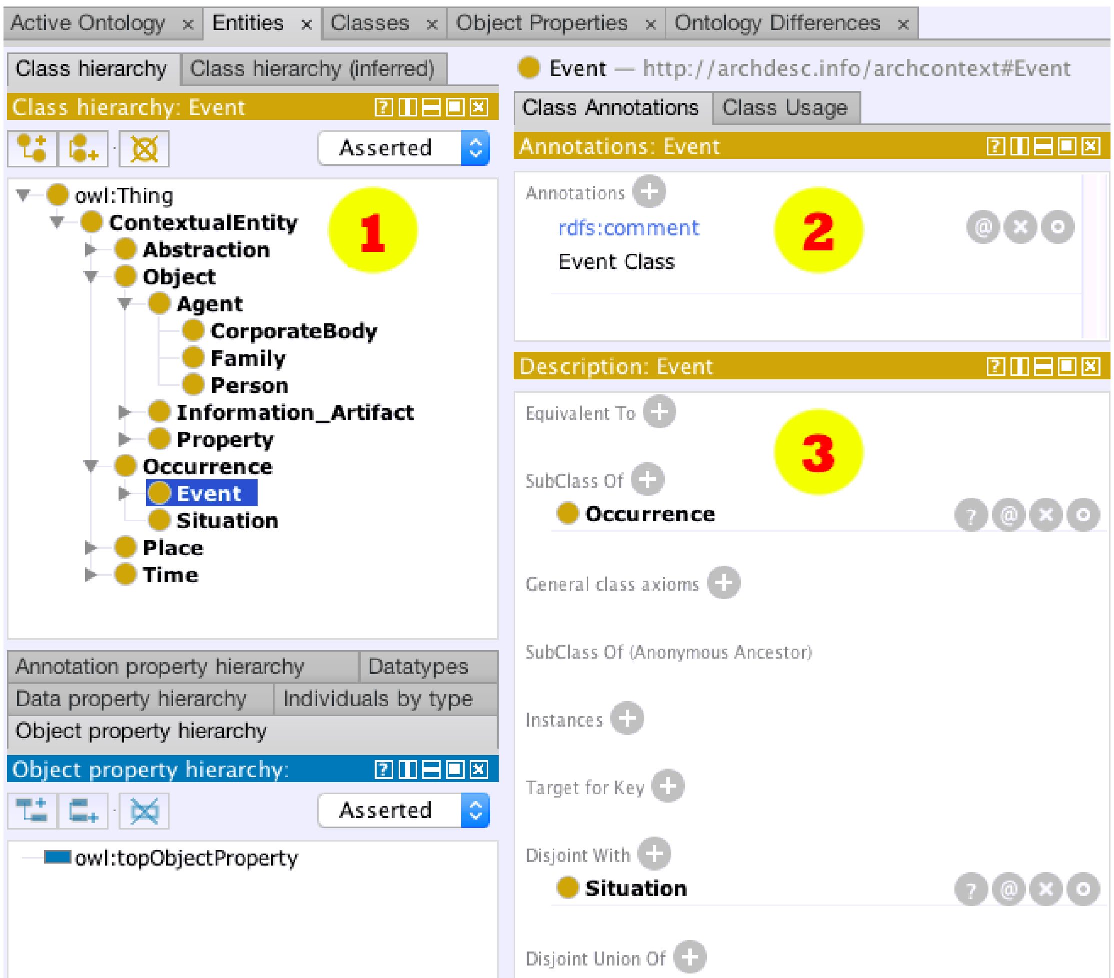


Figure 4.14 Modeling the hierarchical structure and a class *Event*

The class can be exported to different formats such as RDF/XML, OWL/XML, or OWL functional syntax rendering. Table 4.16 shows the class *Event* in OWL/XML format. The header of the ontology consists of lines 1 to 8 in Table 4.16 to define a list of namespaces used in the ontology. Lines 19 to 23 in Table 4.16 correspond to the annotation property “rdfs:comment”, as illustrated in the area 2 of Figure 4.14. Lines 24 to 26 in Table 4.16 define the class *Event*.

```
1 <?xml version="1.0"?>
2 <Ontology xmlns="http://www.w3.org/2002/07/owl#"
3 xml:base="http://archdesc.info/archEvent"
4 xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
5 xmlns:xml="http://www.w3.org/XML/1998/namespace"
6 xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
7 xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
8 ontologyIRI="http://archdesc.info/archEvent">
9 <Prefix name="" IRI="http://archdesc.info/archEvent#"/>
…
15 <Annotation>
16 <AnnotationProperty abbreviatedIRI="rdfs:comment"/>
17 <Literal datatypeIRI="http://www.w3.org/2001/XMLSchema#string">A core ontology
created by Jason Zou for describing archival description. </Literal>
18 </Annotation>
19 <AnnotationAssertion>
20 <AnnotationProperty abbreviatedIRI="rdfs:comment"/>
21 <IRI>#Event</IRI>
22 <Literal datatypeIRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral">Event
Class </Literal>
23 </AnnotationAssertion>
24 <Declaration>
25 <Class IRI="#Event"/>
26 </Declaration>
27 </Ontology>
```

Table 4.16 The class Process in OWL format (partial view)

Task 2: Build the hierarchical structure of classes.

The hierarchical relation between two classes is a sub/super-class (or is-a) relation. For example, *Person* is a sub-class of *Agent* and *Agent* is a super-class of *Person* (*Person* is-a *Agent*).

The sub/super-class (or is-a) hierarchy can be built in two ways: 1) by creating a class as a sub-class of its super-class directly, and 2) by creating a class and then positioning the class into the hierarchical structure. The first way can be done in area 1 of Figure 4.14 through the first icon under “Class hierarchy: Event”, which is for adding sub-classes. For example, the class *Event* (a sub-class of the class *Occurrence*) can be created with the following steps: 1) selecting (clicking the class *Occurrence*); 2) then clicking the “add sub-class” icon; 3) typing in “Event” for the sub-class that is about to be created; and 4) finishing the submission. Then, the class *Event* is created. The second way is to create a class in area 1 and then position the class by using the "*SubClass Of*" in Area 3 to place the class into the hierarchy. After the class is created, Protégé validates the names of classes to prevent wrong characters or duplications.

The disjoint-composition can be specified in the "Description" area of the interface. Figure 4.14 shows that at the bottom of area 1, *Event* is disjoint with *Situation*. The partition-composition and exhaustive-composition need to be incorporated in other ways to formally define the partition and the exhaustive features of the composition.

A class can be further defined by determining whether the class (concept) has equivalent classes (see the *Equivalent To* in area 3 of Figure 4.14); whether this class is associated with instances (see the *Instances* in area 3 of Figure 4.14); and whether the class is disjoint with other classes (see the *Disjoint With* and *Disjoint Union Of* in area 3 of Figure 4.14).

Task 3: Model relations.

OWL properties represent relations. Binary relations are relations between two classes and can be defined in the "Object Properties" tab (the area 1 as illustrated in Figure 4.15). The area 2 is for annotations and usage of the selected property in the area 1. The occurred_at property is chosen in Figure 4.15. The area 3 is for defining the details of a property. In Figure 4.15 the area 3 illustrated the relation is between class *Event* and *Time* of the occurred_at relation. The "Characteristics" area specifies whether the relation is one of the seven kinds of properties: functional, inverse functional, transitive, symmetric, asymmetric, reflexive, and irrelative.

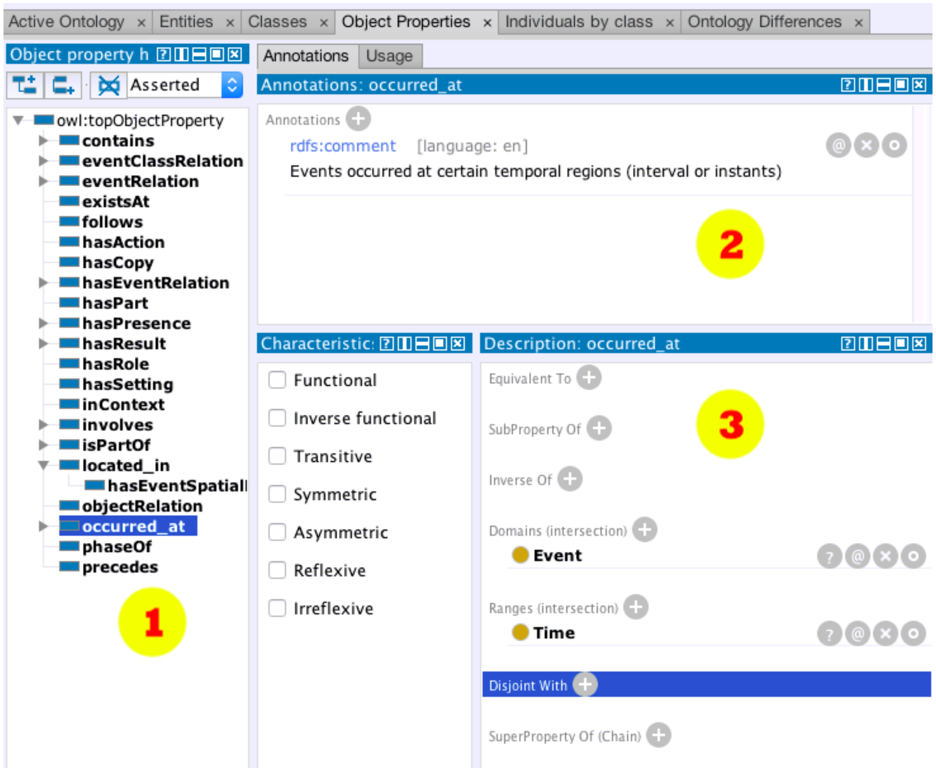


Figure 4.15 The occurred_at relation

A functional property refers to a "property that can have only one unique value y for each instance x" (Bechhofer et al., 2004). For example, a person can only have one biological mother. The relation can be expressed in Table 4.17. Based on the definitions, the instantiation statement implies that :Mary and :Susan are :Woman and they refer to the same person.

Definition	**Instantiation**
FunctionalObjectProperty(:birthMother) ObjectPropertyDomain(:birthMother :Person) ObjectPropertyRange(:birthMother :Woman)	:John :birthMother :Mary, :Susan

Table 4.17 An example of a functional object property

An inverse functional property uniquely identifies the subject (source or domain) in a relation. For instance, the property "social insurance number" is an inverse functional property of an individual: a person only has one social insurance number and the number is unique.

A transitive object property interlinks two individuals A and C, whenever it interlinks A with B and B with C for an individual B. For example, a property hasAncestor links individuals John, Tim, and Mike, as shown in Table 4.18. The statements imply that John hasAncestor Mike.

Definition	**Instantiation**
TransitiveObjectProperty(:hasAncestor) ObjectPropertyDomain(:hasAncestor :Person) ObjectPropertyRange(:hasAncestor : Person)	:John :hasAncestor :Tim :Tim : hasAncestor :Mike

Table 4.18 An example of a transitive object property.

A symmetric object property indicates that the direction of a relation does not matter, meaning the subject (source or domain) and the object (target or range) of a relation can be exchanged with each other. For instance, a property associateWith links two individuals A and B (i.e., A associateWith B and B associateWith A). Either way, the statements hold true.

Although an asymmetric object property connects A with B, it never connects B with A. For example, a property *hasParent* is an asymmetric object property because if John has the parent Mike, then Mike never has the parent John. A reflexive object property "relates everything to itself", while an irrelative object property cannot be related to itself. For instance, a property *parentOf* is an irrelative object property because nobody can be his/her own parent.

Task 4: Model attributes.

Instance attributes "describe the instances of a concept" (Gómez-Pérez, Fernández, & Vicente, 2009, p. 3). Instance attributes can do mapping to DataType properties, another type of property that OWL provides. In Protégé, Data Properties are what connect instances (individuals) with data values. For example, John is an individual of the class *Person*. The *dateOfBirth* datatype property can be used to specify John's date of birth, as in:

```
DataPropertyAssertion(:dateOfBirth :John "2000-01-01"^^xsd:date )
```

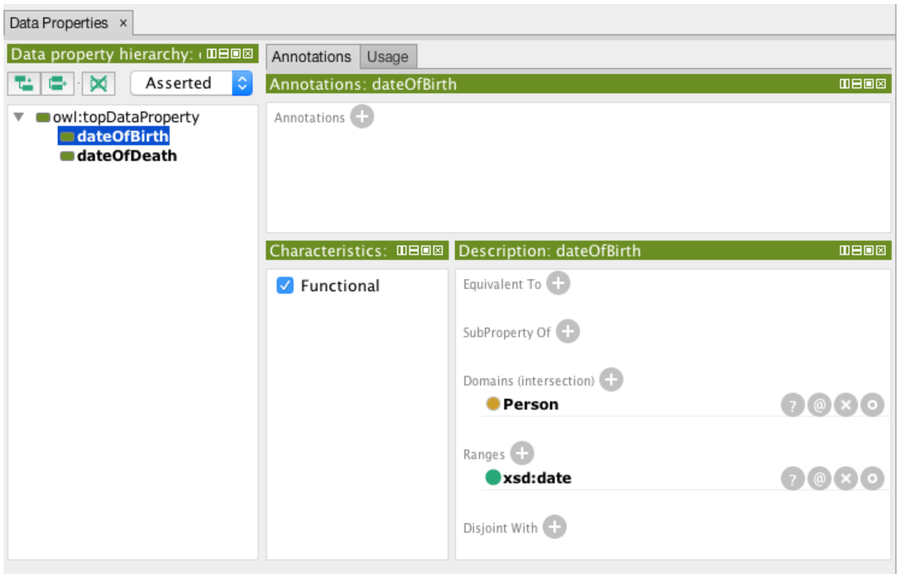


Figure 4.16 The dateOfBirth property

Figure 4.16 shows that the domain of the *dateOfBirth* is a class *Person* and range is xsd:date.

Task 5: Model instances

Instances are called individuals in Protégé. An individual (instance) can be defined in the individual view, which includes annotations, descriptions, and property assertion tabs. The class hierarchy is also for linking instances to classes. Figure 4.17 shows an example of an instance *BirthOfDaleThomson,* which is an instance of *BirthEvent* in the example.

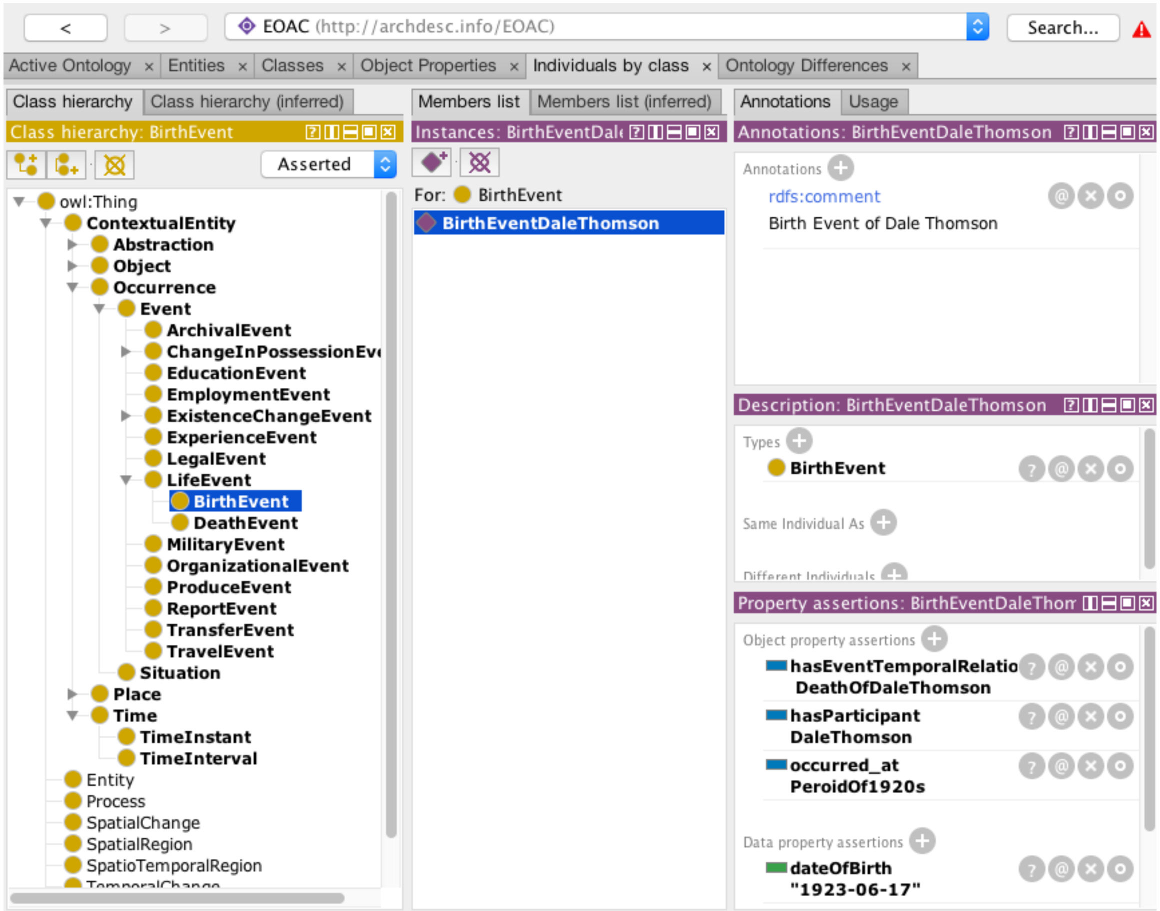


Figure 4.17 A birth event instance

It has an *hasEventTemporalRelation* property which links to another instance *DeathOfDaleThomson* that describes the death event of Dale Thomson. The *hasEventTemporalRelation* can link the two instances. More specifically, if the *before_evt* property is used to replace the *hasEventTemporalRelation* property, it means that the *BirthOfDaleThomson* event happened before the *DeathOfDaleThomson* event. The data property *dateOfBirth* indicates when the *BirthOfDaleThomson* event occurred.

4.2.5. Stage 5. maintenance

The next stage is maintenance. The focus of this maintenance stage is the evaluation of the methontology procedure because the built ontology needs to be verified as to whether the developed ontology meets the defined specifications and whether it contains any errors (e.g., loop definitions). In addition, the documentation of the evaluation facilitates the reusability of the ontology design. Although the ontology needs to be updated and maintained over time, the refinement and maintenance of the ontology construction can also provide insights for future development.

The evaluation of the developed ontology has been made through two measurements. First, in Protégé, it provides a list of reasoners that can be used to check inconsistency problems or loop definitions. Reasoners are based on the description logic to check classes, relations, and restrictions. FaCT++ is a tableaux-based reasoner that provides reasoning services for Protégé which supports the OWL description logics (Tsarkov & Horrocks, 2006).

In Protégé, there are two kinds of structures in class hierarchies: asserted and inferred. The two class hierarchies are in the Classes tab, as shown in Figure 4.18.

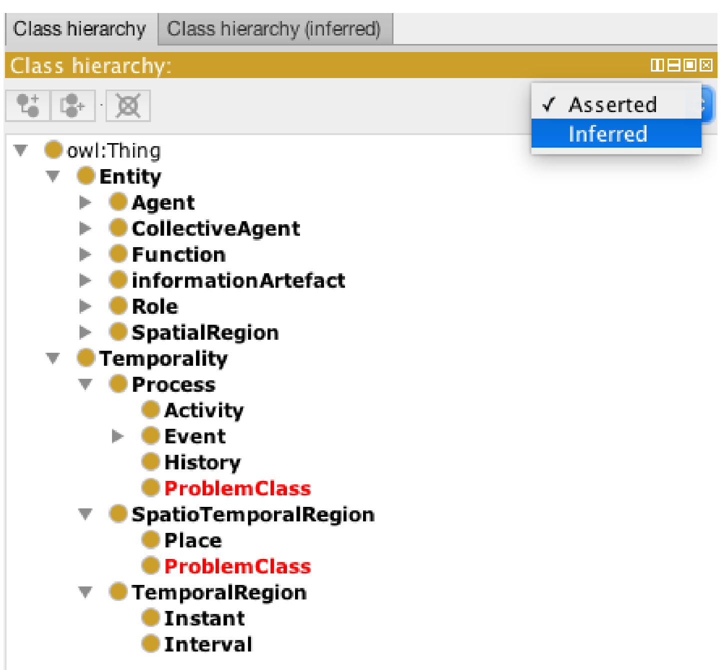


Figure 4.18 The class hierarchies

The asserted class hierarchy is created manually. The inferred class hierarchy is automatically computed by Protégé, based on the asserted class hierarchy. The FaCT++ reasoner can be run through the Reasoner menu, as shown in Figure 4.19.

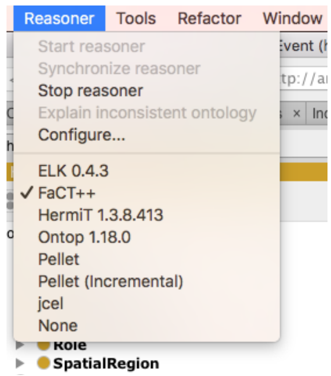


Figure 4.19 The reasoner menu in Protégé

If the FaCT++ reasoner detects inconsistent ontology, it shows inconsistent ontology messages and the "Explain inconsistent ontology" menu item in the reasoner menu between "Stop reasoner" and "Configure…" will be turned on. The explanations of inconsistent ontology are similar to those shown in Figure 4.20 if the menu item is clicked. Figure 4.20 shows an example of the inconsistency of a class *BirthEvent* that is a sub-class of two disjointed classes *LifeEvent* and *EducationEvent.*

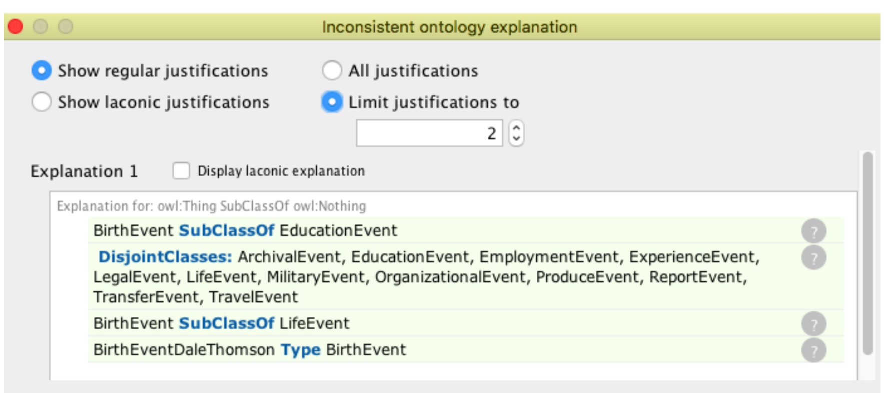


Figure 4.20 Inconsistent ontology explanation

Since the class *LifeEvent* is disjoint with the class *EducationEvent*, this means that one class (e.g., *BirthEvent*) cannot be the sub-class of the two classes at the same time. In this example, the

BirthEvent is not only a sub-class of *LifeEvent* but also a sub-class of *EducationEvent*. This is inconsistent with the disjoint axiom between *LiftEvent* and *EducationEvent*. The inconsistency is detected by the FaCT++ reasoner as shown in Figure 4.20. The FaCT++ was run along with constructing the ontology and inconsistency was dealt with consistently. When the EOAC was completed, the FaCT++ was run, and no error was detected.

Nevertheless, FaCT++ has limitations in which it cannot capture errors related to data properties. Furthermore, to complement the results of FaCT++ as well as to validate whether the ontology meets our specifications, an additional assessment with domain experts and users has been conducted with the focus group study.

In summary, methontology is the method that has been selected for building the ontology. Regarding the construction of an ontology, methontology consists of five stages, including specification, conceptualization, formalization, implementation, and maintenance. Knowledge acquisition activities spread across the entire methontology process, especially in the specification and conceptualization stages. The procedures for the ontology in an archival context have been demonstrated at each stage. The findings of the focus group study will be explained in the next section.

CHAPTER 5. THE FINDINGS OF THE FOCUS GROUP STUDY (STUDY 3)

This chapter presents the findings of the focus group study and answers the second sub-question of the research question (RQ) 2 (RQ. 2.2 Are ontology technologies perceived by users to be appropriate for representing archival contexts?). The first section describes the characteristics of the participants and the procedures for identifying and analyzing emerging themes. The second section presents the findings of the focus group study.

5.1. Background of participants

Participants were asked to answer a questionnaire (see APPENDIX IV. Sample questionnaire). In the first section of the questionnaire, they were asked to answer background information related to their knowledge and experiences in creating and using archival descriptions. While participants of Group 1 were students, those in Group 2 were professional archivists. The backgrounds of the participants in each group are summarized in Table 5.1 and Table 5.2.

Question	Group 1					
	P1	P2	P3	P4	P5	P6
1. What is your affiliation with McGill University?	MLIS Student	MLIS Student	Doctoral Student	Doctoral Student	Doctoral Student	Doctoral Student
2. Have you taken any courses in archival sciences?	4 courses	None	None	None	5 or 6 courses	None
3. Have you taken any metadata related courses?	1 course	None	1 course	2 courses	1 course	None
4. Have you visited archives?	Yes, 5 times	No	No	No	Yes	Yes, 1 time
5. Have you used archival descriptions before?	Yes	No	No	No	Yes	No

Table 5.1 Backgrounds of participants in Group 1

The first question for Group 1 asked participants about their affiliation with McGill University. Participants in Group 1 have knowledge about archives from taking courses on archives (2 participants) or metadata (another 2 participants), and 2 participants did not have any of this knowledge. They were asked about their experience using archival descriptions.

Participants in Group 2 are professionals who have experience with archives or creating archival descriptions. They have knowledge about archives from taking courses on archives (3 participants) or metadata (3 participants). A question on affiliation was removed from Group 2 and instead they were asked whether they created archival descriptions.

Question	Group 2				
	P1	P2	P3	P4	P5
1. Have you taken any courses in archival sciences?	MLIS in archives	MLIS in archives & more than 6 courses	None	3 or 4 courses from Archives Association of Ontario	None
2. Have you taken any metadata related courses?	1 course	2 courses	2 courses	None	None
3. Have you worked in archives?	Yes, working in an archive	Yes, work there	Yes, about one and a half years	Yes, various periods of time	Yes, daily employ-ment
4. Have you used archival descriptions before?	Yes, in my daily work	Yes, every day	Yes	Yes, occasionally	Yes, 1-2 times a week
5. Have you created archival descriptions?	Yes, in my daily work	Yes, part of my work	No	Yes, occasionally	Yes, 2-5 times/week

Table 5.2 Backgrounds of participants in Group 2

In Group 2, three participants currently work in archives, and two participants have worked in archives. All five participants use archives, and most of them (4 out of 5) have experience creating archival descriptions.

5.2. Findings of the focus group questions

To answer research question 2.2 (i.e., Are ontology technologies perceived by users to be appropriate for representing archival contexts?), the questions in the focus group study were designed to ask for participants' opinions or views of the existing archival descriptions and the ontology-based model. Two types of examples were given to participants. For Questions 1 and 2, participants were given with three examples of existing archival descriptions that were randomly selected from the collected sets of Study 2. For Questions 3 and 4, participants were given with the proposed event ontology model that were ontological representations of the three archival descriptions used in Questions 1 and 2 (see APPENDIX V. Three sample archival descriptions for focus group discussions). The focus group questions and the archival description examples are included in APPENDIX VI. Focus group questions. The findings of the focus group studies are explained with each question.

Question 1: Do you think that the current archival descriptions fulfill your needs as a user? Explain the advantages and disadvantages of using the current archival descriptions?

In answering Question 1, Table 5.3 presents responses from participants from the two groups. The second question in Question 1 was divided by advantages and disadvantages.

Emerging Themes for Question 1	Group 1						Group 2					SUM
	P1	P2	P3	P4	P5	P6	P1	P2	P3	P4	P5	
I think that the current archival descriptions fulfill my needs as a user.	√				√		√	√	√		√	6
Advantages												
Archival descriptions are useful.	√				√		√	√			√	5
Archival descriptions are unique.									V			1
Disadvantages												
Archival descriptions lack linkages between related records.			√	√		√			√		√	5

Archival descriptions provide limited information for users.	√				√		√		√			4
Many archival descriptions are not searchable.	√									√	√	3
Archival descriptions are difficult to use due to professional jargon.	√											1

Table 5.3 Responses to question 1 by themes

Specifically, in regards to using archival descriptions, participants indicated six kinds of themes in total (i.e., two kinds of advantages and four kinds of disadvantages). Regarding the advantages, among the six participants who agreed on the first question, five participants mentioned that archival descriptions are useful. The participants remarked:

> "If you know how to use finding aids. …. finding aids are *extremely useful*" (G1-P1).
> "Archival description is actually *very useful* because it does provide the context of the records, the people who created the records "(G1-P5).

Archival descriptions are also addressed as unique by saying "the content is *unique*, [archival descriptions] are the only way in so they are *better than nothing"* (G2-P3).

However, both groups addressed more disadvantages of using the current archival descriptions than advantages. First of all, five participants remarked that archival descriptions lack linkages between related records by stating:

> "linkages are *lacking* between records which make it "difficult to navigate from one record to another" (G2-P3).
> There needs "to be better intermeshing of relation topics.... Sometimes I *can't tell how the related things are* just from the records descriptions unless you actually go in" (G2-P5) because "[we] need to create *a relation* between archival description and creator and the content of records" (G1-P6).

Secondly, four participants addressed that archival descriptions provide only limited information by noting:

> “Archival descriptions are somewhat *superficial* in terms of the depth information they can provide” (G2-P3).
> Archival descriptions are “created in a *very cursory glance* at records” and “provide the information easiest to find” by archivists (G2-P1).
> Furthermore, "Archival records have *not* described *the way*” that the users expect them to be. (G1-P5).
> “[Archivists] *don’t tend to provide a lot of access points* within archives.” (G1-P1).

Thirdly, archival descriptions are not searchable, which is one of the major problems in web environments. For example, G2-P5 remarked:

> "archival records are *still mostly physical records*," and not all digitized archival records (including archival descriptions) were "OCRed and made the content searchable… If this information is available online and in a format like [what] they are used to interacting with on websites and social media platforms, archival description is going to be a lot more accessible to them...My workplace will not be in digital archives" (G2-P5).

There can be failures in searching and locating archival records that are not interactive in web environments.

Lastly, the current archival descriptions are not easy to understand because of professional jargon. Based on the responses to Question 1, participants in both groups seem to perceive that the existing archival descriptions have limitations for users.

In summary, two advantages and four disadvantages of archival descriptions have been identified by the two groups. Six out of eleven participants consider current archival descriptions fulfill their needs. It is worth noting that the six participants include four out of five professionals and only two out of six participants from the novice group. Participants identified two kinds of advantages and four kinds of disadvantages of using archival descriptions. The current archival descriptions seem to have more disadvantages to the user groups than professionals.

Question 2: Do you think the current archival descriptions fulfill your needs as a professional if you are a professional or are assumed to be a professional? Explain the advantages and disadvantages of using the current archival descriptions?

A summary of the responses from participants regarding Question 2 is presented in Table 5.4. The second question was divided according to emerging themes of advantages and disadvantages.

Emerging Themes for Question 2	G1						G2					SUM
	P1	P2	P3	P4	P5	P6	P1	P2	P3	P4	P5	
I think that the current archival descriptions fulfill my needs as a professional.	√						√	√			√	4
Advantages												
Archival descriptions provide meaningful contextual information.							√		√			2
Tools that support current descriptive practices are available.							√					1
Disadvantages												
Archival descriptions cannot keep up with changing context.					√		√	√			√	4
Archivists have limited time to create archival descriptions.	√				√		√	√				4
It is difficult to update existing archival descriptions.					√							1
It is difficult to present multiple creators in one archival record.							√					1

Table 5.4 Responses to question 2 by themes

Based on the responses, four participants out of the 11 individuals agreed that current archival descriptions fulfill their needs as professionals, which is a lower number of participants than Question 1. Participants in both groups also addressed the two kinds of advantages and the four kinds of disadvantages of using current archival descriptions. Regarding the advantages of using current archival descriptions, only two participants raised the first theme by stating that the current archival descriptions provide "meaningful contextual information" of records (G2-P1) and "access points including creators and subject headings" (G2-P3). The second theme was addressed by one participant, who commented there are "tools that support creating and managing archival descriptions" (G2-P1). This aspect is positive with regards to enabling more archival descriptions to be published online than ever before. It is noteworthy that these advantages are only addressed by Group 2.

In the meantime, participants tend to discuss more disadvantages than advantages, as in Question 1. There are two main disadvantages of using the current archival descriptions. Four participants addressed that current archival descriptions cannot keep up with changing contexts by stating that "Places changes… there are merges and de-merges…. Terminology changes and we have to address that street name changes, cities change names, and countries change names" (G1-P5). Similarly, G2-P5 commented:

> "I have worked on archival descriptions related to an old town whose name had been *changed to a new one*. If someone is searching by the new name of the town, they will *never find them*, because the descriptions are listed under the old town" (G2-P5).

The second main disadvantage is addressed by four participants, who said that archivists are "[pressed] for time" and there is no time for a "very deep reading" when they process archival records and create archival descriptions (G2-P1). Within the limited time, archivists have to strike a balance between the "level of description you can do and the amount of time you have" (G2-P2). G1-P1 remarked, "doing [it] even at the series level would be extremely difficult - just *time-consuming* and costly."

Other disadvantages include that current archival practices make it difficult to update the existing archival descriptions and present the multiple creators in one archival record because "a record must belong to one identifiable records' creator", and it is difficult to "recognize the roles of other individuals who have contributed to the records and made up the fonds" (G2-P1).

Along with the responses to Question 1, both groups identified more disadvantages than advantages in using the current archival descriptions as both of users and professionals. The major disadvantages of the current archival descriptions include a lack of linkages between records, providing limited information, being insufficiently searchable, hard to keep up with archival descriptions' changing contexts and limited time constraints for creating them. Participants seem to think that the advantages of current archival descriptions may not yet mitigate the problems they are presently addressing.

In summary, four out of eleven participants believed that archival descriptions fulfill their needs as professionals. Among these four participants, three are from the professional group, which reflects that the current archival descriptions do not seem to satisfy users, as found in Question 1.

Question 3: Do you think that an ontology model, especially the proposed event model, fulfills your needs as a user or professional? Explain the advantages/disadvantages of using the ontology model?

Before asking Question 3, the three sample descriptions in ontology-based models were given to participants in three different kinds of representations: 1) Example 1 presented events in a table; 2) Example 2 presented events in temporal relations; and 3) Example 3 presented events-associated entities in directed graphs (See APPENDIX V. Three sample archival descriptions for focus group discussions). A summary of the responses to Question 3 is presented in Table 5.5.

Emerging Themes for Question 3	Group 1						Group 2					SUM
	P1	P2	P3	P4	P5	P6	P1	P2	P3	P4	P5	
I think that an ontology model, especially a proposed event model, fulfills my needs as a user or professional.	√		√	√		√	√		√		√	7
Advantages												
An ontology-based model adds more access points (i.e., structured information).	√			√	√	√	√	√	√	√	√	9
An ontology model could connect records and reveal relationships.	√		√			√	√	√	√			6
An ontology model disambiguates terms clearly in different contexts.				√	√							2
Disadvantages												
It is time-consuming and labour intensive to apply an ontology model to archival descriptions.	√				√	√	√	√				5

Table 5.5 Responses to question 3 by theme

Regarding a response to Question 3, seven participants out of 11 individuals in the two groups agreed that an ontology model, especially the proposed event model, fulfills their needs; the number of participants has increased in comparison to the responses to Questions 1 and 2. Three kinds of advantages and one disadvantage of using an ontology model are addressed by participants. The primary advantage of using an ontology model is, as nine participants addressed, that applying an ontology model to archival descriptions can provide more access points to archival descriptions. This advantage is apparently agreed upon and closely related to providing more structured information for archival descriptions. One participant commented, “ontology can provide *more access points* to archival descriptions” (G1-P1). Another participant also remarked that applying ontology technology to archival descriptions "adds *an extra layer of information*. You are just giving them more information, more access points" (G1-P5). Describing archives in ontology adds *more structured information* including "setting up semantic relations between topics" (G1-P6). Ontology technologies add "semantic meaning to some entities" (G1-P4).

As for the second advantage, six participants addressed that the adoption of the ontology model to archival descriptions could raise the possibility of establishing inter-connections between records. Ontology technologies "allow you to link related records by categorizing those *types of relations*. If it could be done well, it would strengthen the finding aids" (G2-P1). In the same record, ontology technologies ensure "the *relationships* in the person section are potentially extremely useful when you want to find your way through this kind of historical data and *different relations* between people" (G2-P3). In addition, the proposed event model can be used to make connections and show relationships, as per the comment:

> "Events can be "*related* to each other with the same information (such as the same date, title)…This is just an extra step that we can use to *make a connection*" (G1-P1)
> "All *relations* [that] are represented in graphs are *easier to understand*" (G1-P6)
> "I can see archival description #3 working fine. There are *more keywords and information* you could *interconnect* with other records. In #1 and #2, I could not see more ways to *inter-connect* records with each other….#3 has *more mark-ups* and [is] *less obvious*" (G2-P2).

The third advantage of using the event model is that, since ontology technologies explicitly define semantics, they are useful for clarifying different terms in different contexts. One participant remarked, "one of the advantages of using [an] ontology like the proposed model would be [to] really *disambiguate the meanings* of certain terms" (G1-P5).

As for the disadvantages of using the ontology model, five participants addressed time and cost related concerns by stating that:

> "It would be extremely difficult, just *time-consuming*, *costly* to translate the finding aids using the proposed model" (G1-P1).
> "It is *labour intensive* to do the tagging" (G1-P5).

"Although it will be useful to tag [that] information, it will take *a huge amount of time*" (G2-P1),
"It is *a lot of work* to apply the ontology model to archival descriptions by annotating manually" (G2-P2).

As raised in the disadvantages of Question 2, archivists tend to feel pressed for time once again if the ontology model is used. Furthermore, when "obsolete archival descriptions reside with lots of problems" (G1-P5), this would be an even greater disadvantage.

Consequently, the proposed event ontology model is perceived to be appropriate for fulfilling their needs as both users and professionals. With regards to ontology technology, nine participants tended to have the opinion that it enables them to add more access points and provide more structured information to archival descriptions. However, some participants also have practical concerns about applying the ontology model to existing archival descriptions due to time, labour and resource issues.

In summary, seven out of eleven participants (e.g., four participants from the user group and three participants from the professional group) considered the proposed event model to fulfill their needs. Nine out of eleven participants perceived that an ontology model could bring about more access points.

Question 4: Would you use an ontology model, particularly the proposed event model, at your workplace?

This question intended to ask participants' opinions on whether they would implement and use the ontology model or not. The responses were summarized in Table 5.6 by emergent themes.

Emerging Themes for Question 4	G1						G2					SUM
	P1	P2	P3	P4	P5	P6	P1	P2	P3	P4	P5	
I will use an ontology model, especially the proposed event model, at my workplace.	√	√	√		√	√	√	√	√	√	√	10

Primary Condition											
If there are (semi-automatic or automatic) tools to facilitate applying the model.	√	√	√		√	√	√	√	√	√	√	10
Additional Conditions											
If the event model can meet the high user expectations of applying ontology technology.	√	√				√						3
If the consistency concerns on applying the proposed model are solved.							√	√	√			3
If a workable demo can be provided to show how the ontology, especially the event model, integrates with existing tools.									√		√	2

Table 5.6. Responses to question 4 by themes

Regarding Question 4, ten participants out of 11 individuals from both groups agreed to use the proposed event model at their workplace, which is positive for this study. However, it is conditional. The conditions include the following four aspects: the implementation of workable tools, user expectations, consistency, and the availability of workable demonstration. The primary condition is to have semi-automatic or automatic tools to implement the ontology-based event model. With this condition, ten participants are willing to adopt the proposed event model, which reflects that the facilitated tools are important in adopting the proposed ontology model. In other words, the participants may not use the model without those tools. Thus, a key issue is developing the tools to use the ontology model.

The other three conditions were also addressed by several participants. Three participants in Group 1 focused on meeting the high user expectations of applying ontology technology. Users' expectations include "archival descriptions can be *searched [in] the way they find* today's weather on their cell phone" (G1-P1). Users would like to see a "*Google-like search*" with "better results returned" (G1-P6). It is of interest that concerns about high user expectations were only addressed by participants in Group 1. Three participants in Group 2 were concerned about how the proposed event model could be consistently applied to archival descriptions, which was raised only by professionals in Group 2. Three participants commented,

> "[I]t is hard to say how many relationships you are proposing, and how many relationships people are adopting" (G2-P1).
>
> "[I]f you are diligent and try to remember to tag everything, you still miss things you could have tagged" (G2-P2).
>
> "It is possible that people may tag *differently* for the same thing. Even for myself, I may tag *something in a specific way*, after some time I may tag the thing *another way*" (G2-P3)

As the third condition, two participants (G2-P3 and G2-P5) would like to see a working demo for applying the model to archival descriptions.

The responses of the participants from both groups addressed very practical perspectives to consider in the adoption of the ontology-based event model. It is apparent that their primary condition for adopting the model is to use it as a supporting tool to minimize the time or labour in applying the model. This may be caused by the fact that archival professionals tend to feel the pressure of time in creating archival descriptions and have a large backlog in their archives.

Regarding the overall summary of the focus group studies, participants from the two focus groups responded to the same set of questions with different perspectives. Participants from Group 1 focused more on problems with searching and using the current archival descriptions (i.e., lack of linkages, limited information, insufficient searches, and difficulty of use due to professional jargon). Participants from Group 2 focused more on the problems related to the creation of archival descriptions (i.e., keeping up with changing contexts, limited time to create archival descriptions, difficulty in updating archival descriptions, and difficulty in presenting multiple creators in archival descriptions). However, both groups identified more disadvantages than advantages in the current archival descriptions. Regarding the proposed event ontology model, seven out of eleven participants from both groups perceived that the event ontology model would fulfill their needs with conditions. Ten out of eleven participants were willing to adopt the proposed event ontology model if supporting tools were available.

CHAPTER 6. DISCUSSION

This research has examined what the essential components of archival context are, how the components are related to each other and how they can be represented using ontology technologies. This chapter discusses the major issues and implications drawn from the findings of Chapters 4 and 5. This chapter is organized into four sections. In the first section, the key findings of the research questions are summarized and explained. In the second section, the issues of archival context are explained with regards to the existing knowledge gaps that have been identified in the literature review. In the third section, the issues regarding the focus group study and their implications are discussed. In the fourth section, issues about ontology building procedures are explained.

6.1. Answers to research questions

6.1.1. Research question 1

Research Question (RQ) 1 (i.e., what entities are essential components of archival context?) aims to understand archival context by identifying the essential components of archival context, the characteristics of these components and relationships among these components.

RQ 1.1 What entities are used to provide information on archival context in existing archival descriptions?

Through the content analysis of Study 1, archival descriptions have been categorized as general entities. After the dissection of archival descriptions into sentences and small granular entities, the findings of Study 1 show that archival context consists of ten types of entities: *Agent*, *Artifact*, *Event*, *Feature*, *Function*, *Place*, *Relation*, *Role*, *Situation*, and *Time*. These entities are major ones describing archival context at a high level. Since one entity can belong to only one type, they are mutually exclusive.

The next step is to identify their relationships to reveal the important aspects of entities since the usefulness of entities can be improved when they are linked together.

RQ 1.2 What are the relationships of those entities in representing archival context?

Among the ten entities identified in the results of the RQ 1.1, an entity *relation* is used to link between two entities. Any two of the nine entities (i.e., *Agent*, *Artifact*, *Event*, *Feature*, *Function*, *Place*, *Role*, *Situation*, and *Time*) can be part of a relation (e.g., the *play a role* relation links an *Agent* and a *Role*). In addition, two entities of the same type can be in a relation (e.g., *has a sub-event* relation links an *Event* to another *Event; part of* relation links a *Place* to another *Place*).

In this study, three types of relationships have been identified, including 1) kinship relations (e.g., *son-of*, *daughter-of*, *parent-of*, etc.) that are with *Agents*; 2) social relations (e.g., *colleague-of*, *friend-of*, etc.) that are also with *Agents*; and 3) other relations to link identified entities (mostly between *Agent* and other entities).

6.1.2. Research question 2

Research question (RQ) 2 (i.e., To what extent can ontology technologies be used for representing archival context?) is to explore whether ontology technologies can represent archival context and whether the contextual entities identified in Study 2 and ontology technologies are useful for archival context.

RQ 2.1 What components are necessary for representing archival context when using ontology technologies? How can ontology technologies be used to represent the information on archival context?

The process of applying a methontology approach to Study 2 has been conducted, and the following outcomes were produced: 1) ontology requirement specification; 2) a glossary of terms; 3) an ad-hoc binary relation table; 4) a concept dictionary; and 5) the final ontology model for archival context (see APPENDIX XVI. A complete event ontology for archival context). With the ten contextual entities that have been identified in Study 1, Study 2 explored whether the entities can be classes regarding ontology and what ontological relationships among the entities exist. *ContextualEntity* is a class at the top level in the built ontology. *ContextualEntity* has the direct sub-classes of *Abstraction, Artifact (Object), Feature, Occurrence, Place,* and *Time. Abstraction* is the super-class of *Role* and *Function, Artifact (Object)* is the super-class of *Agent* and *Information_Artifact,* and *Occurrence* is the super-class of *Event* and *Situation.* Since the entity *Relation* describes the relationships between any two of the nine entities, in the Web Ontology Language (OWL), the property can be used for the same purpose, so that *Relation* is replaced with the property in OWL. The *Relation* is further converted into the binary representation in OWL (sub-property of the owl:topObjectProperty). The *Feature* for describing the features (attributes or qualities) of other entities cannot exist without the entities that the *Feature is* describing. Parts of *Feature* are converted into sub-properties of the owl:topDataProperty (e.g., Bill Clinton's first name can be modelled as a DataProperty denoted as DataProperty(Person:first_name, range(xsd:string) and Individual(Person:Bill Clinton, type(Person) value (Person:first_name "Bill"^^xsd:string))) and "Bill" is a type of string). Table 6.1 shows the summary of contextual entities and their hierarchical structure represented in the OWL.

Contextual Entity		**Representation in Ontology (OWL)**
Abstraction		Direct sub-class of *ContextualEntity*
	Role	Direct sub-class of *Abstraction*
	Function	Direct sub-class of *Abstraction*
Artifact		Direct sub-class of *ContextualEntity*
	Agent	Direct sub-class of *Artifact*
	Information Artifact	Direct sub-class of *Artifact*
Feature		Direct sub-class of *ContextualEntity* Direct sub-property of owl:topDataProperty
Occurrence		Direct sub-class of *ContextualEntity*
	Event	Direct sub-class of *Occurrence*
	Situation	Direct sub-class of *Occurrence*

Place		Direct sub-class of *ContextualEntity*
Relation		Sub-property of *owl:topObjectProperty*
Time		Direct sub-class of *ContextualEntity*

Table 6.1 Contextual entities represented in the ontology

In Table 6.1, *Abstraction* and *Occurrence* are classes that have been further developed in the ontology building process in Study 2, which are core entities of the Event Ontology for Archival Context (see APPENDIX XVI. A complete event ontology for archival context). As the contextual entities are formally represented in an ontology, the next step is to explore whether the proposed event ontology model is perceived to be appropriate by users and professionals.

RQ 2.2 Are ontology technologies perceived to be appropriate for representing archival context?

To answer RQ 2.2., focus group studies have been conducted. Participants from the two focus groups responded to the same set of questions from different perspectives. Based on the findings of the focus group study, users and professionals seem to be well aware of the advantages and disadvantages of using the current archival descriptions. It is not surpurising to seen that both groups identified more disadvantages than advantages of existing archival descriptions. Participants of Group 1 addressed problems in using the existing archival descriptions from the perspective of users, such as a lack of linkages, limited information, and that descriptions are not searchable and they are difficult to use because of professional jargons. In the meantime, participants from Group 2 addressed problems related to the creation of archival descriptions from the perspective of professionals, such as being difficult to keep up with changing contexts, limited time available to create archival descriptions, the difficulty in updating archival descriptions and the need to present multiple creators in archival descriptions. Regarding the proposed event ontology model, seven out of eleven participants from both groups seem to perceive that the event ontology model appropriately fulfills their needs. Ten of the eleven participants were willing to adopt the proposed event ontology model on the condition that supporting tools are available. The themes that emerged are integrated into two types of themes: expectations of users regarding archival descriptions, and participants' perceptions of using ontology technologies, specifically the proposed event ontology. The detailed themes and a comparative analysis are discussed in section 6.3, A comparative analysis of focus group findings.

In summary, to provide information on archival context, ten entities (i.e., *Agent*, *Artifact*, *Event*, *Feature*, *Function*, *Place*, *Relation*, *Role*, *Situation*, and *Time*) and three types of relations (i.e., kinship, social relation, and relations to link *Agent* with other entities) have been identified. By using the methontology, the ten entities and relations have been conceptualized and formalized into an event-oriented ontology. In the focus group study, the two focus groups identified the advantages and disadvantages of using existing archival descriptions and the event-oriented ontology from users' and professionals' perspectives. To the participants of both groups, the event-oriented ontology is perceived to be appropriate to represent archival context.

Based on these findings, major issues are discussed in the next three sections in the following order: 1) issues related to archival context at the three levels; 2) a comparative analysis of focus group studies; and 3) issues on ontology building procedures.

6.2. Archival context at the three levels

As the contextual entities and their relationships are identified and formalized in an ontology in Studies 1 and 2, modeling the archival context in the conceptual models for presenting archival context are explained in the following sections.

6.2.1. Conceptual models for archival context

To examine archival context, with regards to the definitions of records, some researchers consider records as "persistent representations of occurrents" (Yeo, 2007, p. 136) and see records as social actions (Foscarini, 2013). This view indicates that a record is the result of social actions and that understanding an activity, action, or event that occurs is central to recognizing records and archival context. In addition, other researchers address that archival context has been extended to social and cultural aspects of context (Bastian & Alexander, 2009; Lemieux, 2014, 2015; Nesmith, 2006; Piggott, 2012; Yeo, 2007). In line with this view, Bastian (2003) states that the community is seen both "as a record-creating entity and a memory frame that contextualizes the records" (p. 3). They see that as context is indefinite (Duff & Harris, 2002) and limitless (Yeo, 2012); archival context is multi-faceted and complex. So, a single level view of archival context is not sufficient to understand all facets of the complexity of archival context.

To understand the cultural, social and technological aspects of context, cultural-Historical Activity Theory (CHAT) has been introduced (Yamagata-Lynch, 2010). In this theory, an activity system is the main tool for modelling context and consists of *Tool, Subject, Rules, Community of significant others, Division of labor,* and *Object* (Engeström, 1987) among which creator-related components of archival context are considered to be more important than other components.

To model archival context for digital collections, several conceptual models have been developed, as has been explained in Chapter 2. Literature Review, Section 2.1.3, including the Context Model (Lee, 2011), the Europeana Data Model (Europeana Project, 2013), the Finnish Conceptual Model for Archival Description (National Archives of Finland, 2003), the Records in Contexts (Experts Group on Archival Description, 2016), and the Spanish Model (Comisión de Normas Españolas de Descripción Archivística, 2012). These models each have advantages and disadvantages. Regarding contextual entities, the Context Model divides contextual entities into nine classes in a flat structure (Lee, 2011). The Europeana Data Model consists of five contextual entities and is a general model for cultural heritage communities (Europeana Project, 2013). The Finnish model is a combination model in which archival contextual entities are mixed with the concepts from library communities (National Archives of Finland, 2003). The Spanish Model was built on the Record's model (McKemmish et al., 2006) and recognized archival reality, conceptual world, and the world of representations that are different from other models (Comisión de Normas Españolas de Descripción Archivística, 2012). Among these conceptual models, the Records in Contexts (RiC) model is worth attention as a comparison with this study because it has recently been introduced by the International Council of Archives (2016) and it is a more specific and updated model for representing archival descriptions. The RiC model consists of 14 classes (i.e., *Record, Record Component, Record Set, Agent, Occupation, Position, Function, Function (Abstract), Activity, Mandate, Documentary Form, Date, Place,* and *Concept/Thing*) and 792 relations (Experts Group on Archival Description, 2016). This model seems to capture the most prominent contextual entities of archival context. Recalling that this study focuses on context-related entities, the relevant properties and their relations to the RiC model are compared in detail with the proposed model of this study. Table 6.2 illustrates a comparison between the RiC model and the model proposed in this study according to entity.

Entity	**This Study**	**Records in Contexts (RiC)**	**Close Match**	**Partial Match**
Agent	Agent	Agent	√	
Artifact	Artifact	Concept/Thing		√
Community	Community subclass of Agent			
Event	Event	Activity	√	
Feature	Feature			
Function	Function	Function,	√	

		Function (Abstract)		
Mandate	Rule	Mandate	√	
Place	Place	Place	√	
Relation	Relation	Object Properties		√
Role	Role	Occupation, Position		√ √
Situation	Situation			
Time	Time	Date	√	
Tool	Artifact			
		Documentary Form		
		Record, Record Component, Record Set		
Sum			6	4

Table 6.2 Comparison of contextual entities in the two models

The two models have six entities that are closely matched and four entities that are partially matched, as illustrated in Table 6.2. Since the model in this study has a focus on archival context, some entities *(e.g., Record, Record Component,* and *Record Set)* of the RiC model do not have corresponding entities.

The RiC model has several limitations. Its entities do not align with any upper-level ontology and RiC-Ontology is not available. Although the RiC model *relation* is not defined as an entity, the 792 properties defined in the model can be considered as types of relations. Unfortunately, the rationale for selecting 792 properties is not explicitly stated. Because two properties (such as *is associated with* and *was associated with*) are semantically identical in the RiC model, if they are counted as one property *associate with*, the number of properties can be reduced to half.

Furthermore, the rationale of selecting 14 entities in a RiC model is not explicitly clear in some entities. For example, entities *Concept* and *Thing* are combined and their definition is not completely clear (e.g. "[a]ny idea or notion, material thing, or event or occurrence that can be associated with, or in some cases be the subject of, other entities" (Experts Group on Archival Description, 2016, p. 18). In addition, two entities, *Function and Function (Abstract),* have no detailed explanation for using the differences between the two entities. The RiC model does not have an explicit view to connect 14 entities. Entities *Community, Tool,* and *Rule* are not included. Most importantly, the RiC model does not include a clear explanation regarding how to relate itself to archival theory.

In comparison, this study produced the Event Ontology for Archival Context (EOAC). It is built on the Basic Formal Ontology (BFO) that is a widely used and mature upper ontology. This model enables entities, properties, and relationships to align with those of other ontologies, an aspect that ensures the interoperability of the EOAC. The entities of this study were identified through the content analysis and further conceptualized via explicitly defined procedures of a methontology. The proposed model identified and focused on three kinds of relationships: kinship relation, social relation, and other general relations (e.g., *is_part, has_member,* and *associate_with*). Contextual entities link to each other through identified relations. For instance, an entity *Event/Situation* connects to another *Event/Situation.* In this way, the contextual entities and relations can be shown in the immediate creation context as well as social, cultural and technological contexts, which enable the linkages to weave a complex web of relations and provide meaningful information to users at multiple levels.

The next sections explain the multiple views of archival context at the three levels (i.e., micro, meso and macro levels) to better understand the contextual entities in archival context.

6.2.2. Archival context at the micro-level

Ten contextual entities are identified and formalized in the proposed ontology, as has been explained in Section 6.1. Answers to research questions, the nine entities (i.e., *Agent*, *Artifact*, *Event*, *Feature*, *Function*, *Place*, *Role*, *Situation*, and *Time* except *for Relation)* and the relationships between them have been identified. Based on the entities and relationships, the detailed views of the archival context of the ontology are examined in this section. Since one relationship connects to two entities, all relationships between the entities are illustrated in Figure 6.1.

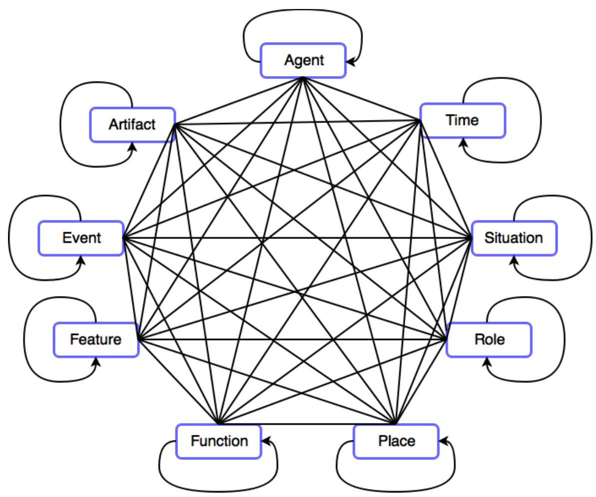


Figure 6.1 The complete relationship between entities

Among all of these relationships, some relationships are more meaningful than others. For example, the *play a role* relationship between *Agent* and *Role* may be treated as a more important relationship rather than a stable relationship with other entities. In most situations, a certain period of the relationship between *Place* and *Time* is stable (e.g., Toronto is a place today and will be a place tomorrow). Furthermore, when archivists recognize archival context, not every relationship in an actual context is identifiable based on documented records. As explained in the identified relationships in Study 1, relationships are built more to *Agent*, which is agreeable with the Spanish Model (Comisión de Normas Españolas de Descripción Archivística, 2012). Among all of the relationships of contextual entities given in Figure 6.1, the next Figure 6.2 shows the common relationships between contextual entities in solid lines rather than dotted lines.

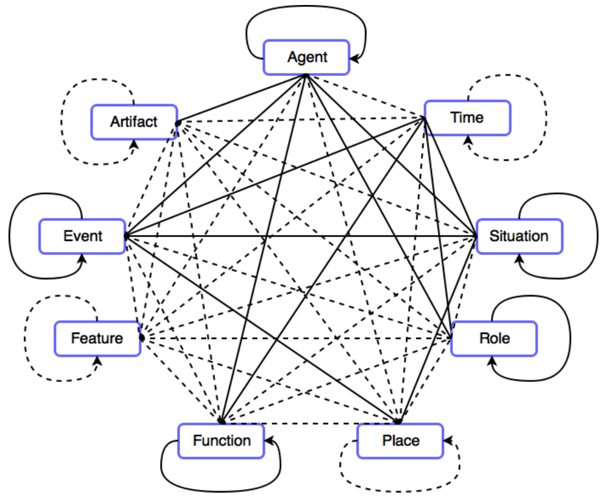


Figure 6.2 Relationships between contextual entities in solid lines

As is illustrated in Figure 6.2, as more relationships that link to *Agent, Event, Situation,* and *Time* have more solid lines than other entities, these relationships are considered to be more common. Among the four entities, *Event* is seen as the core entity because event glues the contextual entities together and, based on the view that records are considered representations of activities (Lemieux, 2014; Yeo, 2007), event is considered to be the most important entity in this study.

To illustrate the event-centered relationships between the identified entities of the EOAC ontology, Figure 6.3 shows the detailed entities and their relationships in this study. In the right side of Figure 6.3, *Event/Situation* is the central point that directly links to *Agent* 2 (as a participant or patient of an *Event/Situation*)*, Time* (when the *Event/Situation* occurred)*, Place* (where the *Event/Situation* occurred)*,* and *Feature* (what the attributes/qualities of the *Event/Situation* are)*.* An *Agent* 1 (as a creator) plays a *Role* in the creation of *Artifacts* when performing a *Function*. Similarly, Figure 6.3 also shows other relationships such as *has a feature* to link *Artifact* and *Feature*. Additional entities can be further inserted to show more relationships between *Role* and *Function.*

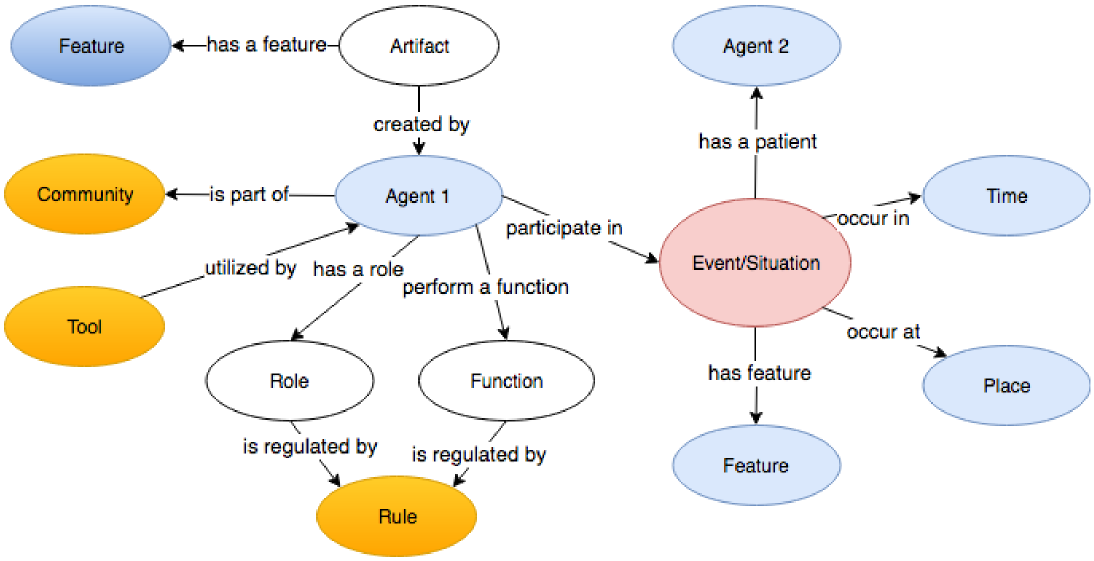


Figure 6.3 A static view of archival context at the micro-level

In addition, *Community, Tool,* and *Rule* (marked in yellow colour) are included in this figure. To present entities and their relationships when building an ontology, there are two kinds of views: 1) a static view and 2) a fluid view. As a static view is a way to show a specific activity of an agent. It is an important step toward modeling entities and relations at the micro-level. Figure 6.4 presents the entities and relationship that are identified in this study around an entity *Event/Situation.* This is the static view of archival context through an event-centric model. A fluid view connects static views into a whole that present all related events, entities, and relations together. A fluid view (aka. span) presents related events, entities and relations in a period of time, while a static view (aka. snap) shows a snapshot of them at a given time. Therefore, static views are the fundamental blocks of a fluid view. Around the main entity *Event/Situation*, the chain of connected *Event/Situations* is the core of fluid view. Figure 6.4 shows the three *Event/Situations* linked in the chain and the three *Agents* are not connected with each other. The agents may be the same person or different people who have participated in different events/situations.

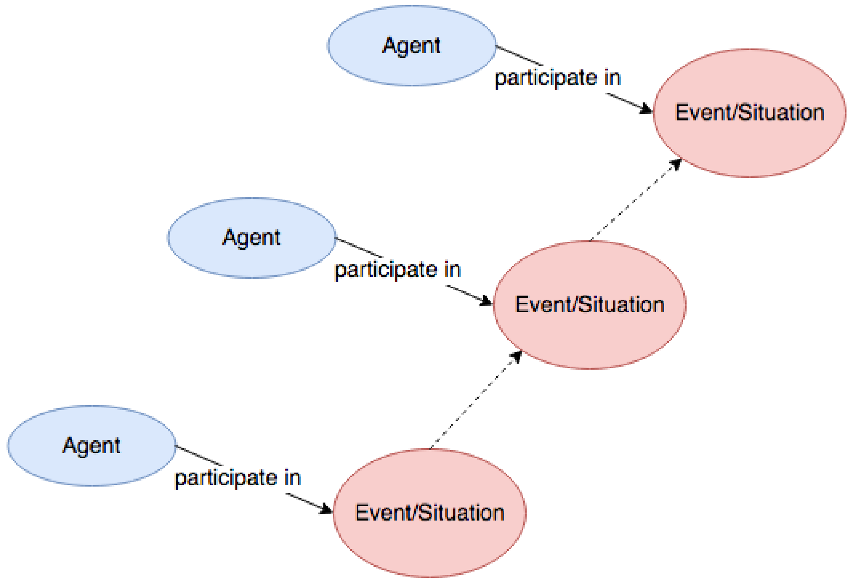


Figure 6.4 A fluid view of archival context at the micro-level

While each *Event/Situation* corresponds to each *Agent* as a static view of Figure 6.3, at the same time, one *Event/Situation* links to another *Event/Situation* to which each *Agent* links in Figure 6.4. The fluid view can represent a history of an agent, such as the agent who performed certain functions, played roles in communities, and created artifacts (e.g., archival records). The most common relations between *Event/Situations* are temporal relations (e.g., before, after, etc.). Through linked events, situations, and related entities, the dynamics of archival context can be captured.

Events and situations are related, and one event may be connected to two situations. As explained in Chapter 4. Findings, 4.2.3.3. Occurrence, two relations (i.e., *precede* and *follow*) are used for connecting an event and two situations. According to Allen (1983)'s 13 types of relationships between two events, *Event/Situation* can be chained up by temporal relationships. In Study 1, 13 types of events are identified, including *ArchivalEvent, BioEvent, BusinessEvent, CommunicationEvent, EducationEvent, LegalEvent, LifeEvent, MilitaryEvent, NaturalEvent, OrganizationalEvent, OtherSocialvent, PoliticalEvent, TravelEvent.* Each type of event can be further divided into smaller events. For example, a type of *BioEvent* can be divided into *BirthEvent* and D*eathEvent*. In addition, events can be grouped by themes (e.g. *LifeEvent, OrganizationalEvent, ArchivalEvent,* etc.) and grouped by types of changes (e.g., *ChangeInPossessionEvent, ExistenceChangeEvent, ExperienceEvent, ProduceEvent,* and *TransferEvent).* These groupings are not mutually exclusive so they can co-exist.

As the EOAC ontology aligns with the upper ontology BFO and an event is a property of a spatiotemporal region (Lewis, 1986), events in this study are carried out within the spatiotemporal regions which means that *Time* class is a sub-class of *TemporalRegion* and *Place* is a sub-class of *SpatioTemporalRegion.*

6.2.3. Archival context at the meso-level

The meso-level of archival context captures the essence of records and explains context in between the macro and micro levels. It is based on the Recordkeeping Metadata model (McKemmish et al., 2006), which consists of three kinds of contexts: 1) creation context, 2) description context, and 3) usage context. Each context has a triad relation with three components, respectively. Each context operates within a macro-level of archival context that may not overlap with each other. Figure 6.5 displays the three kinds of contexts with detailed components at the meso-level.

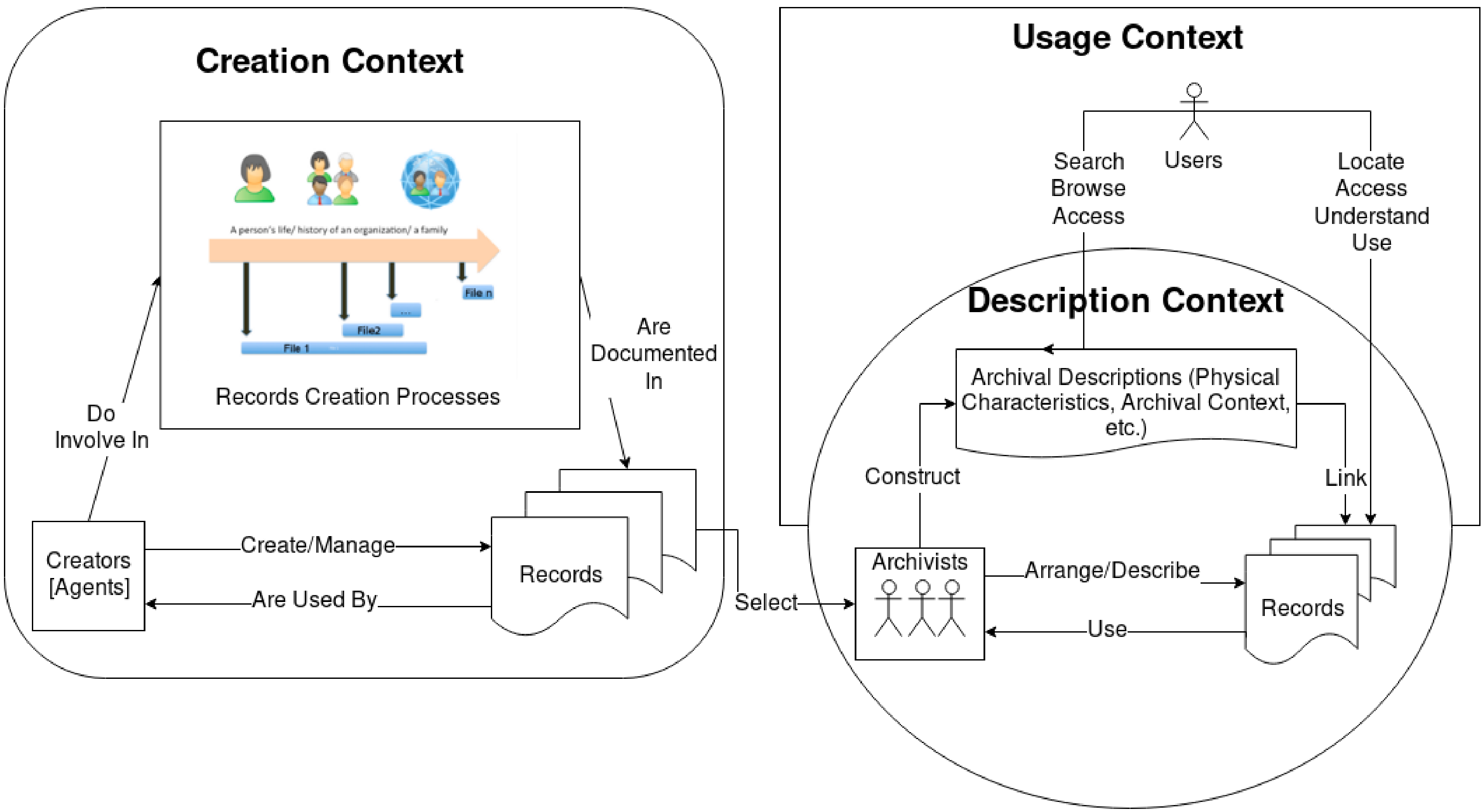


Figure 6.5 A view of archival context at the meso-level based on the recordkeeping model (McKemmish et al., 2006)

As illustrated in Figure 6.5, the creation context consists of creators, records, and record creation processes as a triad; the description context consists of archivists, records, and archival descriptions as a triad; the usage context consists of users, archival descriptions, and records as a triad. In the creation context, creators create and manage records that are used by the creators. In other words, the creators (i.e., agents including individuals, families, and organizations) are not involved in records creation processes that are documented in records. Thus, in the creation context, records creation processes are the most important part.

To classify context, Berkhofer (1995) categorized context into actual context, documentary context, and interpretive context. Actual context refers to "the reality of the world and the experiences of the people in it" (Berkhofer, 1995, p. 20). Documentary context indicates "the documentary or other artifactual sources" (Berkhofer, 1995, p. 20). Interpretive context refers to "the represented context synthesized from the evidential or documentary context" (Berkhofer, 1995, p. 20). Creation context exists independently and is situated in the "real world (i.e., archival reality)" (Comisión de Normas Españolas de Descripción Archivística, 2012, p. 8). Among Berkhofer's categories, the creation context corresponds to an actual context. The creation context can also be recorded in the documentary context. In addition, since interpreted context is a part of archival representation (Yakel, 2003) or a world of representations (Comisión de Normas Españolas de Descripción Archivística, 2012), archival description documents may reflect the archivists' interpretation of the actual and documentary contexts during the records creation processes. Thus, it can also be an interpreted context by Berkhofer's categories. In integrating the three contexts, the interpreted context may reflect creation context that is produced in the description context in archival descriptions.

The description context is the context in which archivists select, arrange, and describe records. The description context is what archival representation occurs (Yakel, 2003). When archivists are reordering, interpreting, and creating surrogates to represent actual archival materials, archivists "need a deconstruction of the contexts they are trying to describe" (Cook, 2001, p. 32). The deconstruction is not about "destroying in endless relativist critiques, but about constructing, about seeing a new and imagining what is possible when the platitudes and ideologies are removed" (Cook, 2001, p. 22). In the course of the inquiry, reading, and analysis, archivists construct an interpreted context through creating and linking descriptions closely to the appraisal reports that justify why the records are in the archives and being described (Piché, 1998). In this way, archivists reveal what the description context includes, what affects the decision-making, why the records are arranged in a certain way and what is included in the descriptions.

When producing the description context, the interpretations by archivists have played an important role in representing the actual context (i.e., archival reality) because the actual context affects a user's search, location, and understanding of archival records through descriptions. By producing the description context, archivists represent the description context explicitly in a meaningful sense, which makes it easier for users to understand and use the records. Thus, looking at archival context and the detailed description context through an ontology helps represent archival context in a formalized conceptual model, understand archival context explicitly and provide a solid foundation for building better archival description systems.

The usage context is a context in which users search, locate, access, and use archival records. The usage context is not necessarily the same as the description or creation contexts. The interactions occur between users and archival descriptions, users and records, and users and archivists. It is important that users reside in the usage context and archivists work in the description context. In digital settings, an archival description representation system is the extra layer between users and archival descriptions. Through the system, users search and access archival descriptions and retrieve archival context that is developed by archivists to reflect the creation context of the creators; users locate and access archival records associated with the archival descriptions, and they try to understand the records through documented records and constructed context. Although technology makes easier to disseminate archival descriptions, it brings barriers to access and navigate descriptions (Chapman 2010; Daniel & Yakel, 2010; Gerhold, 2013).

The three contexts are interconnected through records. Figure 6.5 shows that archivists select records created in the creation context and then arrange the records. Archivists act as the channel to connect the creation context and the description context through arranagement and description activities. Meanwhile, users may browse/search archival descriptions that were created by archivists and then access records that are selected and arrranged by archivists. Figure 6.5 is a simplified view within which users, creators, and archivists are separately functioned. In other words, they do not play mutual roles. However, in the records continuum model (McKemmish, Reed, & Piggott, 2005), it is important that the creation, description, and usage contexts are intertwined. The three contexts can map to the three aspects (i.e., create, process, and use) of records. To better understand and analyze archival context at the higher level, the meso-level of archival context should be considered.

6.2.4. Archival context at the macro-level

Since archival context has multiple aspects, as explained in *6.2.1. Conceptual models for modelling archival context*, more than just the internal parts of records, archival context supports “the social and technical processes of the records’ inscription” (Nesmith, 1999, p. 146) and the “context of cultural expression” (Rowat, 1993, p. 199). In light of this view, context can be considered “functional context” and “regulatory context” (Comisión de Normas Españolas de Descripción Archivística, 2012, p. 26).

The macro-level of archival context is intended to deal with context at the highest level, which is associated with social, cultural, and technological aspects of context. Within the macro level of archival context, the meso-level of archival context operates under where the micro level of archival context operates. More specifically, the macro-level of archival context consists of laws, rules, and regulations, technological tools, and broader communities that are associated with the records or records creators. For example, the creation of records is under certain rules or regulations. The records creation processes of the micro-level include tools that have been utilized in the process or communities to which the creator(s) or the agent(s) of the records belong. In the previous Figure 6.3, *Community, Tool,* and *Rule* are embodied in the static view and the entities of archival context are at the macro-level as well.

In cultural historical activity theory (CHAT), the unit of analysis is an activity (Nardi, 1996) and an activity is usually "carried out not in isolation but in collaboration with others within the community … in a context or a situation where there are rules and conditions that determine the way activity is performed" (Mwanza, 2002, p. 75). In CHAT, human activities are situated in communities and within a broader context to understand activities. The model of human activity is represented by the "activity system" (see Figure 2.3, The model of an activity system) that includes the following components: *Subject, Object, Community, Tools, Rules,* and *Division of Labour* (Yamagata-Lynch, 2010). To operationalize the activity theory, Mwanza (2002) developed the eight-step-model to gather and analyze data from an activity theory perspective. The eight-step-model includes: 1) activity of interest, 2) objective, 3) subjects, 4) tools, 5) rules &

regulations, 6) division of labour, 7) community, and 8) outcome (p. 128). Based on Mwanza's model, Table 6.3 presents the modified procedure for identifying high-level entities in the eight steps.

Steps to be identified	**Questions to ask**
Step 1. Event	What sort of event is the record about?
Step 2. Place and time	Where and when is the event taking place?
Step 3. Agent	Who is involved in carrying out the event?
Step 4. Artifact	By what means are the agents carrying out the event?
Step 5. Rule	Are there any cultural norms and rules governing the performance of the event?
Step 6. Role	What roles do the agents carry out? How are the roles organized?
Step 7. Community	What is the environment in which the event is carried out?
Step 8. Outcome	What is the desired outcome of the event except for the records?

Table 6.3 The procedure for identifying entities at the macro-level

As this procedure is the way in which archivists construct context from documented records, the constructed context is considered a representation and interpretation of the actual context in which the records are created. In this procedure, event is the central point linking entities together. Among the eight-step procedure, Steps 2, 4, 5, and 7 are more important than the others in representing the high level of archival context.

The modified procedure helps archivists identify context when appraising archival records. When executing the procedure first, there is no particular order through the remaining steps of the procedure. So, the procedure can be applied iteratively when appraising and describing archival records.

Step 1 is to identify what the record is about and what events it is related to. For example, the events and social activities may be the focus of a social activist’s fonds. The interactions between the activist and other individuals or organizations may be keys to understanding who is the activist. Thus, this step identifies the boundaries of the components involved.

Step 2 is to determine the place and time the event happens since each event occurs in a certain spatial and temporal environment. The two aspects situate an event in a specific setting.

Step 3 is to identify who is involved in the event. It is the step to identify related individuals, organizations, and other communities.

Step 4 is to identify what types of tools have been used in the event. Artifacts (or tools) can be either material or conceptual, such as language, protocols, scientific methods and models, and other forms of cultural artifacts (e.g., computers or hammers).

Step 5 is to ascertain what cultural and social aspects are directly related to the event. The rules regulate the agent's actions and relations with other participants in the event. For example, an entity *Rule* indicates social, cultural, and regulatory rules.

Step 6 is to determine what roles the agent carried out. Creators perform different roles in various activities (events). For example, an agent professor played multiple roles (e.g., planner, leader, researcher, etc.) to establish and manage a research lab.

Step 7 is to situate the event in its surrounding environment. The community consists of the people who share an interest with agents and involvement with the same event when agents are involved in the selected event with other agents.

Step 8 is to identify accomplishments or achievements of the events. In particular, the direct outcome of the events (e.g., transactions, documents) needs to be identified. In this way, the eight-step procedure can relate entities to records.

Conducting the eight-step procedure makes it possible to identify entities' *event, place, time, agent, artifact (e.g., tool, outcome), role, community, rule* and their relations more clearly and explicitly. In particular, this procedure allows the entities *tool, role,* and *community* to be expanded by further linking to other entities, as can be observed in Figure 6.3. The advantage of using the procedure is identifying archival context at the macro level, where the archival context at the lower levels can be situated.

In synthesis, the conceptual model of this study presents multiple aspects of archival context at the three levels: 1) static and fluid views at the micro-level; 2) three different kinds of contexts (i.e., creation, description, and usage contexts) at the meso-level; and 3) an eight-step procedure to identify archival context at the macro-level. In the procedure of building the event ontology for archival context (EOAC), contextual entities and the relationships that link entities together are identified at each level to examine the complexity of archival context. The micro-level approach, through the static and fluid views of archival context, can dissect archival context into entities and relationships. The meso-level approach can present a comprehensive view of archival context and reveal what aspects of archival context are involved in each creation, description, and usage context, respectively. The macro-level approach expands to show the broad context related to social, technological, and cultural aspects of external environments in which archival records are created. Thus, a comprehensive view of the conceptual framework is possible to better understand the complex aspects of archival context, which is illustrated in Figure 6.6.

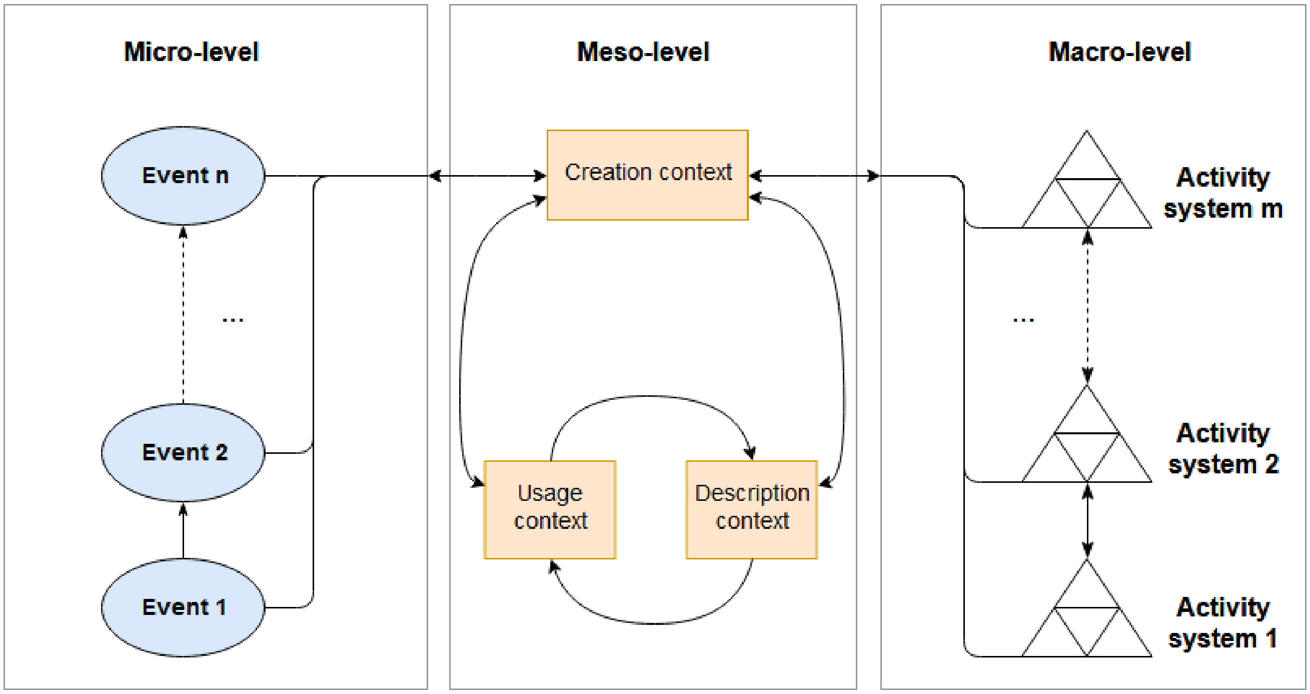


Figure 6.6 The three-level conceptual model for archival context

Figure 6.6 shows a three-level conceptual model for archival context. Each level has its own area of focus. At the micro-level, each oval shape can be depicted as a static view of archival context. Each oval is connected as a fluid view of archival context, which is associated with a creation context of the meso-level. At the meso-level, archival contexts are divided into three kinds: creation, description, and usage contexts. Although the three contexts look separate, they can overlap. At the macro-level, a series of activity systems are interconnected. The components of each activity system can be identified by using the eight-step procedure. The outcome of one activity system may become the precondition of a next activity system (Yamagata-Lynch, 2010) and the six components of each activity system may overlap. The components of each activity system act as a whole that provide the cultural, social, and technological aspects of the creation context for understanding archival context.

The creation context at the meso-level is the core that is associated with a series of activity systems at the macro-level but also can be dissected into a series of static views at the micro-level.

The three-level conceptual model for archival context enables presenting a comprehensive view of archival context with components and their interactions at each level. It also allows us to delineate the details from the specific to the general in a more explicit and elaborate manner. This model is beneficial by providing an operational model for archivists to capture the core entities of archival context, emphasizing the differences between contexts at each level and helping archivists focus on a specific context. This model also provides a feasible way to identify and focus on six components in each activity system that are the building blocks of broader contexts. Altogether, the examination of archival context was possible at the three levels (i.e., micro-, meso-, and macro-levels) both meticulously and comprehensively for investigating the complex nature of archival context.

6.2.5. A use case for understanding the three-level conceputal model for archival context

In order to understand the proposed three-level conceptual model for archival context, the Madeleine Parent fonds (McGill University Archives, 2018) is used as an examplar case. The model is applied from the macro-level to the meso-level and then to the micro-level, with the two activity systems identified in this example.

At the macro-level, the modified eight-step procedure can help identify the detailed components of activity systems. Regarding the years when Madeleine Parent was at McGill, one activity system can be identified by following the eight-step procedure.

Steps to be identified	**Questions to ask**	**Identified Components**
Step 1. Event	What sort of event is the record about?	Study and be an activist
Step 2. Place and time	Where and when is the event taking place?	Montreal, 1937-1940
Step 3. Agent	Who is involved in carrying out the event?	Madeleine Parent, Department of Sociology

Step 4. Artifact	By what means are the agents carrying out the event?	English, French, Books, Sociology Lecture notes
Step 5. Rule	Are there any cultural norms and rules governing the performance of the event?	University regulations, laws and regulations regarding student associations
Step 6. Role	What roles do the agents carry out? How are the roles organized?	Student, activist
Step 7. Community	What is the environment in which the event is carried out?	McGill University, the Canadian Student Assembly, Socialist, Communist influences
Step 8. Outcome	What is the desired outcome of the event except for the records?	B.A degree, campaign experience, leadership skills, collective actions

Table 6.4 Entities of an activity system about Madeleine Parent from 1937 to 1940

Regarding her activities in later years, the same procedure can be used to identify the components of activity systems.

Steps to be identified	**Questions to ask**	**Identified Components**
Step 1. Event	What sort of event is the record about?	Work
Step 2. Place and time	Where and when is the event taking place?	Montreal, 1943-1946
Step 3. Agent	Who is involved in carrying out the event?	Madeleine Parent, Kent Rowley
Step 4. Artifact	By what means are the agents carrying out the event?	Leaflets, Bulletin, Correspondence, notebook, typewriter
Step 5. Rule	Are there any cultural norms and rules governing the performance of the event?	laws and regulations regarding unions
Step 6. Role	What roles do the agents carry out? How are the roles organized?	Union organizer, activist
Step 7. Community	What is the environment in which the event is carried out?	The United Textile Workers' Union of America in Quebec
Step 8. Outcome	What is the desired outcome of the event except for the records?	A strong union

Table 6.5 Entities of an activity system about Madeleine Pareent from 1943 to 1946

Figure 6.7 depicts how the two activity systems are interconnected. One activity system is to identify and link the components that Madeleine Parent was a student at McGill University. The other activity system is to capture her activities as a union organizer in the United Textile Workers' Union of America in Quebec.

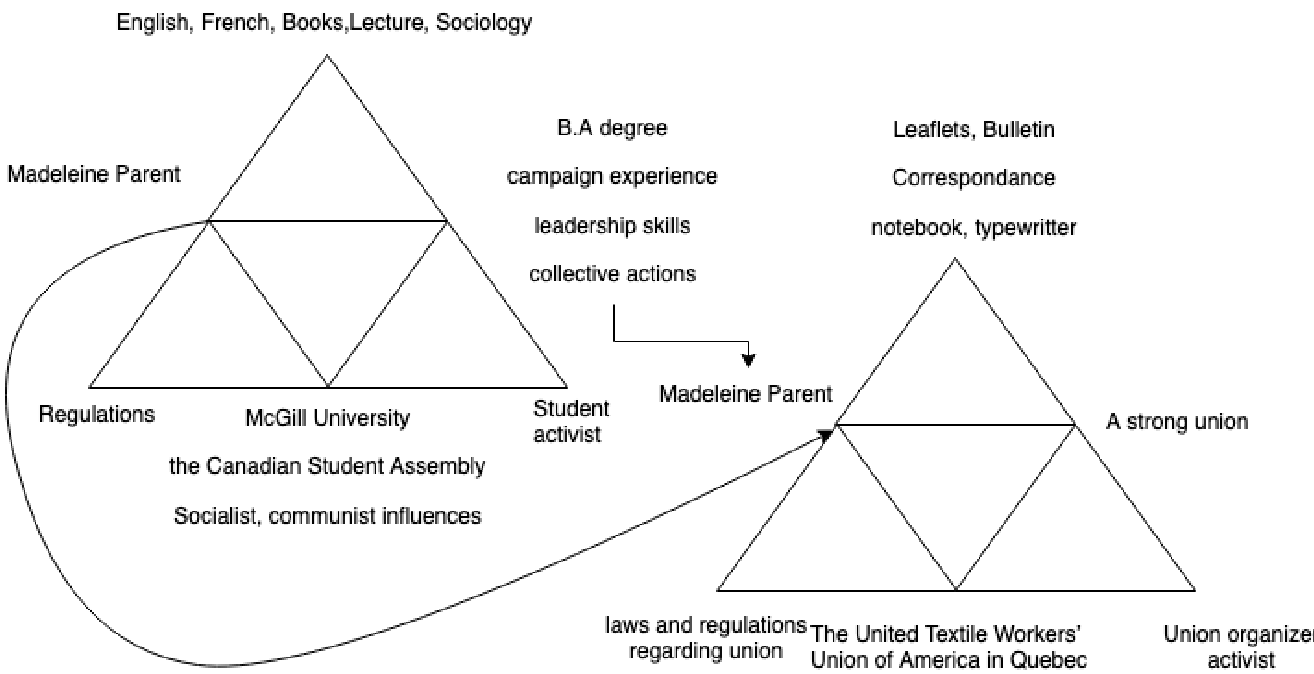


Figure 6.7. An example of two interconnected activity systems

In Figure 6.7, the connections between the two activity systems are clearly displyed, but the temporal and spatio-temporal aspects are not included at the macro-level.

To show how the major components of each event are at the micro-level, links go back to the macro-level through community, rule, and role. Figure 6.8 shows two connected events (an education event and a compaign event).

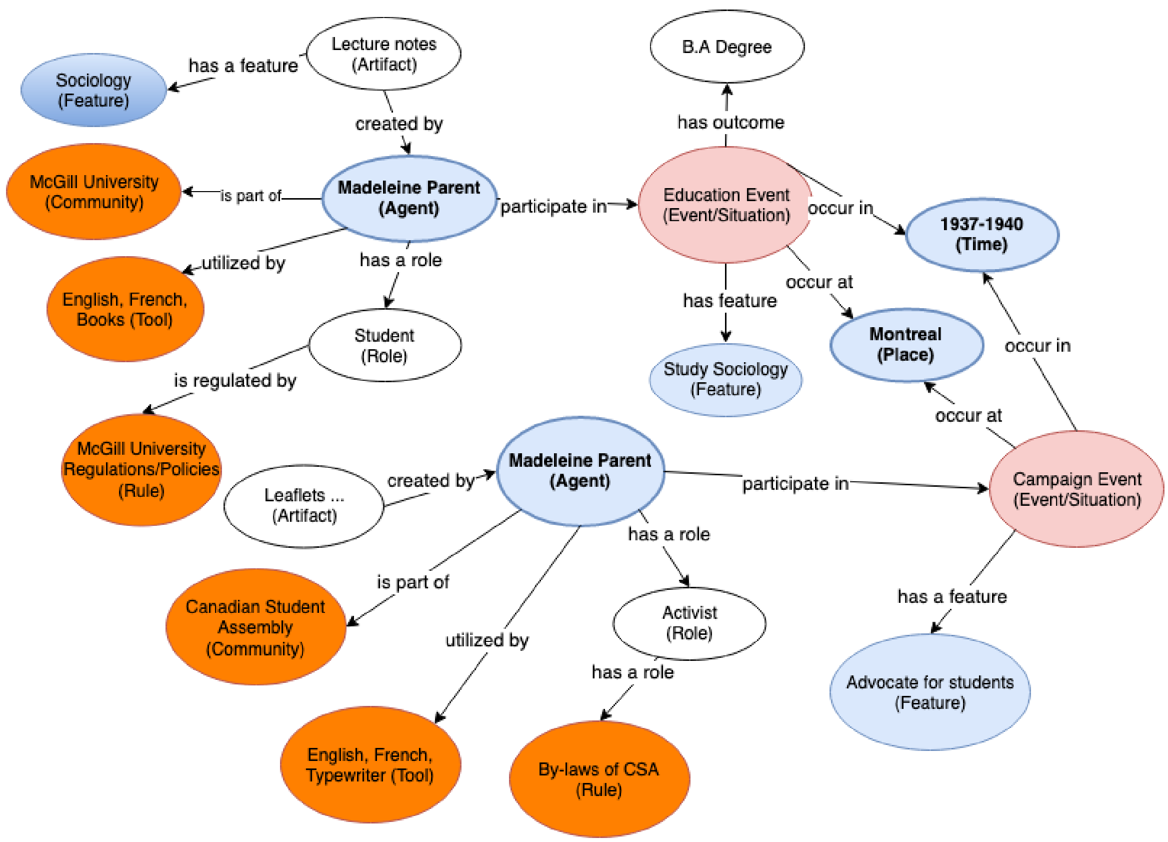


Figure 6.8 Two connected events

The two events were carried out by the same person (i.e., Madeleine Parent) depicted in blue ovals in bold (Madeleine Parent (Agent)). In addition, the two events overlapped in temporal and spatio-temporal regions (depicted as two light blue ovals: 1937-1940 (Time) and Montreal (Place)). Community, Tool, and Rule are part of other entities associated with the activity system on the left side of Figure 6.7. Although this is a simplified view of two connected events, if such connections are established, any event in ovals in Figure 6.8 can lead to the entire figure. This can be helpful for discovering hidden connections.

In summary, based on the identified contextual entities of Study 1 (in Section 6.1. Answers to Research Questions), a domain-driven ontology EOAC model was developed with a focus on explicitly representing archival context. Archival context consists of ten individual entities (i.e., *Agent*, *Artifact*, *Event*, *Feature*, *Function*, *Place*, *Relation*, *Role*, *Situation*, and *Time*) and higher-level entities, such as *Rule, Tool,* and *Community*. A comprehensive view of archival context was developed to examine archival context at the three levels (i.e., micro-, meso-, and macro-levels) in detail.

6.3. A comparative analysis of the focus group findings

RQ 2.2 aimed to explore whether the proposed event ontology model is perceived to be appropriate for both users and professionals. As identified in Chapter 5. The Findings of the Focus Group Study (Study 3), the two groups had different foci. This section presents a comparative analysis of the responses of the two groups regarding the addressed advantages and disadvantages of archival descriptions and the proposed ontology model.

6.3.1. Advantages of archival descriptions – users vs. professionals

The advantages of archival descriptions have been addressed by the two groups. Users remarked upon two kinds of advantages of the existing archival descriptions:

1) User – Adv. 1. Archival descriptions are useful.
2) User – Adv. 2. Archival descriptions are unique.

Professionals are also responded with two kinds of advantages regarding the existing archival descriptions:

1) Prof. – Adv. 1. Archival descriptions provide meaningful context information.
2) Prof. – Adv. 2. Tools that support current descriptions practices are available.

From users' perspectives, archival descriptions are unique (User – Adv. 1) and useful (User – Adv. 2.). In the meantime, the professional group emphasized that archival descriptions provide meaningful context information (Prof. – Adv. 1.), which is resonant with the results of existing studies by addressing that archival descriptions are important (Duff, Craig, & Cherry, 2004) and archival descriptions are used steadily by historians (Anderson, 2004; Tibbo, 2003). Historians value primary sources as the most important materials for their research and use them frequently (Tibbo, 2003) with online tools provided by archives to meet their needs. Tibbo's study (2003) revealed that historians value archival descriptions and 90% of the studied historians used print archival descriptions and 60% of them used electronic archival descriptions. After 2010, it was reported that archival descriptions are still the primary sources for historians to find information about archival records (Rhee, 2012) because there are no other tools to replace archival descriptions that are used for describing unique archival records.

Since then, more tools have been developed, such as Archivists Toolkit (Archivists' Toolkit project, 2009), ArchivesSpace (ArchivesSpace, 2018), and AtoM (Artefactual Systems Inc., 2015) to support current description practices as Prof. Adv. 2. addressed. For example, AtoM is "a web-based, open source application for the standards-based archival description" (Artefactual Systems Inc., 2015, para. 1). AtoM fully supports ISAD (G) (General International Standard Archival Description) and provides search functionality to retrieve archival descriptions. In addition, archivists can use this tool to convert archival description records to EAD (Encoded Archival Description) format or export to an EAD XML format. Tools like AtoM make it easier to create archival descriptions and easily publish them on the Web.

The professional groups know that tools can save time in the creation and publication of archival descriptions. They also understand that existing archival descriptions describe context information and provide valuable information for users. They tend to value the practicality of the use of supporting tools in creating archival descriptions.

6.3.2. Disadvantages of archival descriptions - users vs. professionals

From the users' perspectives, the following disadvantages of existing archival descriptions are addressed:

1) User – Disadv. 1. Archival descriptions lack linkages between related records.
2) User – Disadv. 2. Archival descriptions provide limited information for users.
3) User – Disadv. 3. Many archival descriptions are not searchable.
4) User – Disadv. 4. Archival descriptions are difficult to use due to professional jargon.

Users are not likely satisfied with the current archival descriptions because the linkages between related records are not sufficient (User – Disadv. 1); limited information is given in the archival descriptions (User – Disadv. 2); archival descriptions are not sufficiently searchable (User – Disadv. 3); and professional jargon in archival descriptions make it difficult for users to use (User – Disadv. 4).

A study by Daines & Nimer (2011) recognized four kinds of major problems in existing archival descriptions (i.e., hierarchical display, item-level access, text-heavy blocks, and terminology). These four major problems are similar to the disadvantages of archival descriptions described in this study. As users found that archival descriptions lack linkages between related records (User – Disadv. 1.), a recent study also remarked that a traditional approach would have missed including "links and relationships with other important archives" (Keneley, Potter, West, Cobbin, & Chang, 2016, p. 95). In addition, archival descriptions are created in different record-holding institutions (silos), and it is difficult to link to related records and maintain the links across many institutions. Regarding the links of records, Zhang and Mauney (2013) stressed that an archival record could only become "fully comprehensible when the record has been associated with the […] social context responsible for its creation, […] and other records to which it relates" (p. 175). User – Disadv. 2. indicates that users want more information to be included in archival descriptions. Although one of the purposes of producing archival descriptions is to present contextual and structural information in archival records, archival descriptions tend to provide information at the surface level only. Item-level information in current archival descriptions is often so limited (User – Disadv. 2) that the connection between archival descriptions and item-level information is sometimes missing (Zhang & Mauney, 2013). For example, item-level information is often stored in content management systems such as CONTENTdm (OCLC, 2018). The segregation of archival descriptions and item-level information present an "accumulating discoverability problem" for both users and archivists (Woolcott, Payant, & Skindelien, 2016). Thus, if the additional linkages of records are given to archival descriptions, this enables a search processes to be "easier and [offer] more rapid access to information" (Keneley et al., 2016, p. 94).

User – Disadv. 3 indicates archival descriptions need to be more searchable. It is well-known that on the Web, users tend to search archival descriptions the same way they search a topic on Google or via other portals. They expect archival descriptions to be more searchable because non-searchable archival descriptions make it difficult to find what they want to search for on the Web. Zhou's study (2007) supports the concept that the search functions of archival description systems yield a lower number of retrieval results from archival descriptions than users expect. Since archival institutions tend to make digitized or digital images of archival records available online, archives should publish records in a fully text-searchable way, including archival descriptions and records.

User – Disadv. 4 addresses that professional jargon should not be used if possible because many users of archives find "archival jargon difficult to decipher" (Bahde, 2017, p. 485). As Yakel (2003) stated, this is rooted in the problem of a mismatch between the representations of archivists and users, respectively. In the three-level conceptual model for archival context, users operate in the usage context that may not overlap with the description context, where archivists reside. As Cox (2007) pointed out, an important aspect is that archivists create archival descriptions using the language and manner they are most comfortable with, rather than descriptions that users may seek when using archives. It may be easy and quick for archivists to create archival descriptions with large blocks of text, without proper linkages, and without proper context for users to understand. However, it is a burden for users when they experience “difficulty following large blocks of text in online finding aids” (Bahde, 2017, p. 485).

Gilliland-Swetland (1998) noticed that an understanding of how users navigate and interact with interfaces was missing in the late 1990s. Since then, some studies have been conducted to identify the ways in which users understand and have raised issues (Rhee, 2015; Yakel, 2003; Walton, 2017). To solve the problems that have arisen, several attempts have been made with different approaches, such as the design of user-centered archival information systems (Daines & Nimer, 2011), user participation (Green & Lampron, 2017), the use of emerging technologies and services such as Flickr (Kalfatovic, Kapsalis, Spiess, Van Camp, & Edson, 2008), Open Street Map (Joy, Keane, & Corrigan, 2017), data visualization (Bahde, 2017), a linked data approach (Gracy, 2015) and others. However, because the problems may be rooted in how the archival descriptions can be presented, the EOAC and a comprehensive approach by the three-level conceptual model of this study can be ways to solve the problems raised in meeting users' expectations.

In comparison to users, professionals have identified the following disadvantages of current archival descriptions:

1) Prof. – Disadv. 1. Archival descriptions cannot keep up with changing contexts.
2) Prof. – Disadv. 2. Archivists have limited time to create archival descriptions.
3) Prof. – Disadv. 3. It is difficult to update existing archival descriptions.
4) Prof. – Disadv. 4. It is difficult to present multiple creators in one archival record.

The first disadvantage Prof. – Disadv. 1. (Archival descriptions cannot keep up with changing contexts) may be brought about by several factors, such as a lack of proper tools, limited staff, limited training, complex archival description rules, backlogs, etc. Zhou's study (2007) showed that users expect to find archival descriptions and use archival records through user-friendly interfaces, Google-like search portals, and integrated social media tools. However, the adoption of new technologies in archives tends to be slow due to institutional reasons (e.g., lack of funding, resources, staff, etc.). For example, the adoption of EAD in the United States is slow, with only 42% of the survey respondents using it (Yakel & Kim, 2005). Several researchers addressed that archival information systems do not deliver their expectations sufficiently (Coats, 2004; Rhee, 2015; Zhou, 2007). Thus, the current archival descriptions do not contain changing contexts sufficiently for professionals and users.

Prof. – Disadv. 2. (Archivists have limited time to create archival descriptions) shows that archivists have limited time to create archival descriptions, which is commonly accepted in archives. This disadvantage is related to several factors as well. Backlogs are common problems in archives (Panitch, 2001; Spiro, 2009). Since the creation of archival descriptions is costly and time-consuming, large backlogs hurt the archival profession, which causes a chain of problems, such as blocking access to records (Greene & Meissner, 2005), cursory archival descriptions (Cox, 2007), and requests for more storage space (Ham, 1984),

Prof. – Disadv. 3. (It is difficult to update existing archival descriptions) Implies that archivists are in a difficult situation with the creation, updating, and management of existing archival descriptions. In their limited time, archivists struggle to find resources, deal with scanning paper archival descriptions, and convert different formats to the proper format (such as PDF/A, EAD) in existing archival descriptions. Updating existing archival descriptions is another burden for archivists.

Prof. – Disadv. 4. (It is difficult to present multiple creators in one archival record**)** is resonant with existing studies regarding the provenance problem (Hurley, 1995, 2005; Nesmith, 2006, 2006; Sweeney, 2008). The principle of provenance is to keep records as they were created by the original creator. It is difficult to have multiple creators in one archival fonds since a fonds is an abstract concept rather than a physical one. To solve this issue, the inclusion of linkages among the records can be suggested to connect each creator to his/her contributions. When archivists apply archival description rules to records in practice, this aspect may not be easy for archivists, either.

In comparison to the disadvantages addressed from the users' perspectives, these disadvantages from professionals are correlated with each other. The advantages and disadvantages of archival descriptions from the two groups can be mapped with each other. The mappings are displayed in Figure 6.9.

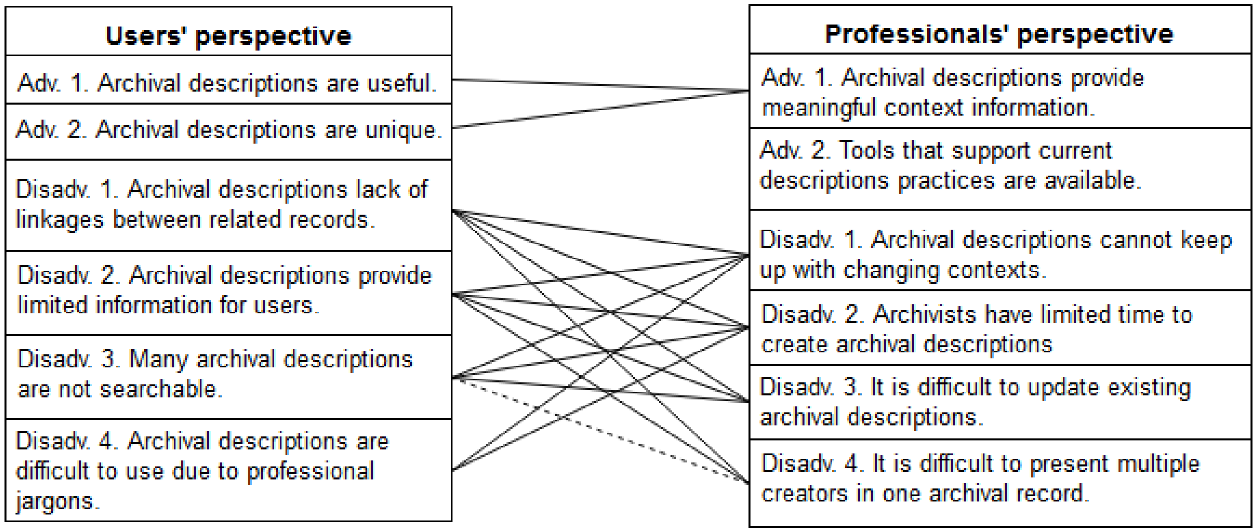


Figure 6.9 Mapping between the advantages and disadvantages in archival descriptions

Each group has identified four kinds of disadvantages of archival descriptions. While the disadvantages identified by professionals are more like problems related to the creation of archival descriptions, those from the users' perspectives are problems related to the use of archival descriptions. In other words, the identified disadvantages from the professionals' perspectives can be correlated with the disadvantages identified by users.

As shown in Figure 6.9, User – Adv. 1 (archival descriptions are useful) and User – Adv. 2 (archival descriptions are unique) are closely related to Prof. – Adv. 1 **(**archival descriptions provide meaningful context information**)** so they directly (denoted by solid lines) correspond**.**

Prof. – Disadv. 1 (archival descriptions cannot keep up with changing contexts) is directly related to all four kinds of disadvantages for users (User – Disadv. 1-4). Users expect archival descriptions to link to related files (item-level) and records within the holding institution and beyond (User – Disadv. 1). Users want more information (User – Disadv. 2) but unfortunately, archivists with limited time, resources, and support fail to meet users' expectations. Adams (2007) addressed that archivists cannot provide such services to entirely meet users' expectations. Users expect archival descriptions and content to be searchable (User – Disadv. 3). Archivists have not moved forward enough to make them all searchable yet. Users have difficulties in understanding professional jargon (User – Disadv. 4). Thus, four direct lines with all of the disadvantages experienced by users can be mapped with archivists' current situations, which implies that archivists should be more user-centered when creating archival descriptions.

Prof. – Disadv. 2 (archivists have limited time to create archival descriptions) is also directly associated with four kinds of disadvantages to users (i.e., User – Disadv. 1-4). When archivists create archival descriptions, including the linkages within and beyond the records, it is very time-consuming, and they have a lack of time, staffing, and resources, as has been mentioned. This may in fact create all of the disadvantages of archival descriptions that have been indicated by users. Regarding professional jargon, as Cox (2007) claimed, archivists create archival descriptions using languages and a manner that they are most comfortable with; it is convenient for archivists to use professional jargon in archival descriptions and it may even save time.

Prof. – Disadv. 3 (it is difficult to update existing archival descriptions) is directly related to the three kinds of disadvantages for users (User – Disadv. 1-3). User – Disadv. 1. (Archival descriptions lack linkages between related records) has an impact on updating archival descriptions. Because of limited linkages, related records without linkages are often ignored. As a result, users have fewer chances to find the related records (User – Disadv. 3.), resulting in User – Disadv. 2. (Archival descriptions provide limited information to users). This situation makes it labour intensive and time-consuming for archivists to update no-search archival descriptions because archivists need to put more effort and spend more time locating and updating archival descriptions.

Prof. – Disadv. 4 (it is difficult to present multiple creators in one archival record) is directly related to two kinds of disadvantages to users (i.e., User – Disadv. 1-2). Prof. – Disadv. 4. has an impact on User – Disadv. 2. (Archival descriptions provide limited information for users). Prof. – Disadv. 4. may affect archivists who are reluctant to create links to connect related records (User – Disadv. 1.). In addition, User – Disadv. 3 (Many archival descriptions are not searchable) is indirectly (denoted by a dotted line) related to Prof. – Disadv. 4.

In Figure 6.9, the mapping of issues shows that the addressed disadvantages of professionals are relating to users' concerns. We see that users tend to have high expectations regarding archival descriptions and current archival descriptions do not meet the needs of users.

6.3.3. Disadvantages of archival descriptions vs. advantages of ontology models

Regarding ontology technologies, participants have perceived that the ontology model, specifically the proposed event-based model, is appropriate. As for the use of the ontology model, participants addressed the following advantages that the proposed event model brings:

Ontology – Adv. 1. An ontology model adds more access points (i.e., structured information).

Ontology – Adv. 2. An ontology model could connect records and reveal relationships.

Ontology – Adv. 3. An ontology model defines terms clearly in different contexts.

These three kinds of advantages identified by participants are the benefits of using ontology and applications, which is agreeable to several researchers as well (Berners-Lee et al., 2001; Breitman, Casanova, & Truszkowski, 2007; Gangemi & Mika, 2003; Hitzler et al., 2009; Hyvönen, 2012). In Ontology – Adv. 1, an ontology model adds a formal layer (specification) of information to archival records through which more access points can be added. Regarding retrieval, the lexical information of the layer and ontological components are rich and empowered (Ranwez et al., 2013).

Ontology – Adv. 2 emphasizes that the ontology model can further connect records with other records and reveal their relationships. By using the Web Ontology Language (OWL), the three basic components of an ontology *Classes, Properties (object* and *datatype*), and *Individuals* are used (Hitzler et al., 2012). The task of *Classes* and *Instances* link individuals to the *classes,* while *object properties* connect related individuals (Hitzler et al., 2009). In an ontology model, people and related entities can be linked together within and beyond archival records and then their relationships are added.

Ontology – Adv. 3 stresses that vague expressions can be eliminated. If a property is related to two entities, this means that the two entities are related and their relation is denoted by the property. For example, if the property is *has_wife* with a domain (*Man*) and a range (*Woman*), the statement <*Sydney has_wife Wendy*> implies that *Sydney* is an instance of class *Man* and *Wendy* is an instance of class *Woman. Sydney* here is differentiated from the city of *Sydney.* This type of information makes the meanings of entities more explicit and clearer, which helps to improve the accuracy rate of information search and retrieval.

In addition, as a domain ontology is also a shared conceptualization of a domain (Borst, 1997), the conceptualization is to build a type of model of a domain (Guarino, 1998) and the built ontology can be reused as a shared consensus of a domain model. Thus, the built ontology model can be further adopted to be used and disseminated in other archives as well.

These three advantages of the ontology model can be full or partial solutions to mitigate the disadvantages identified by the two groups of users and professionals; these relations are displayed in Figure 6.7 through a mapping table.

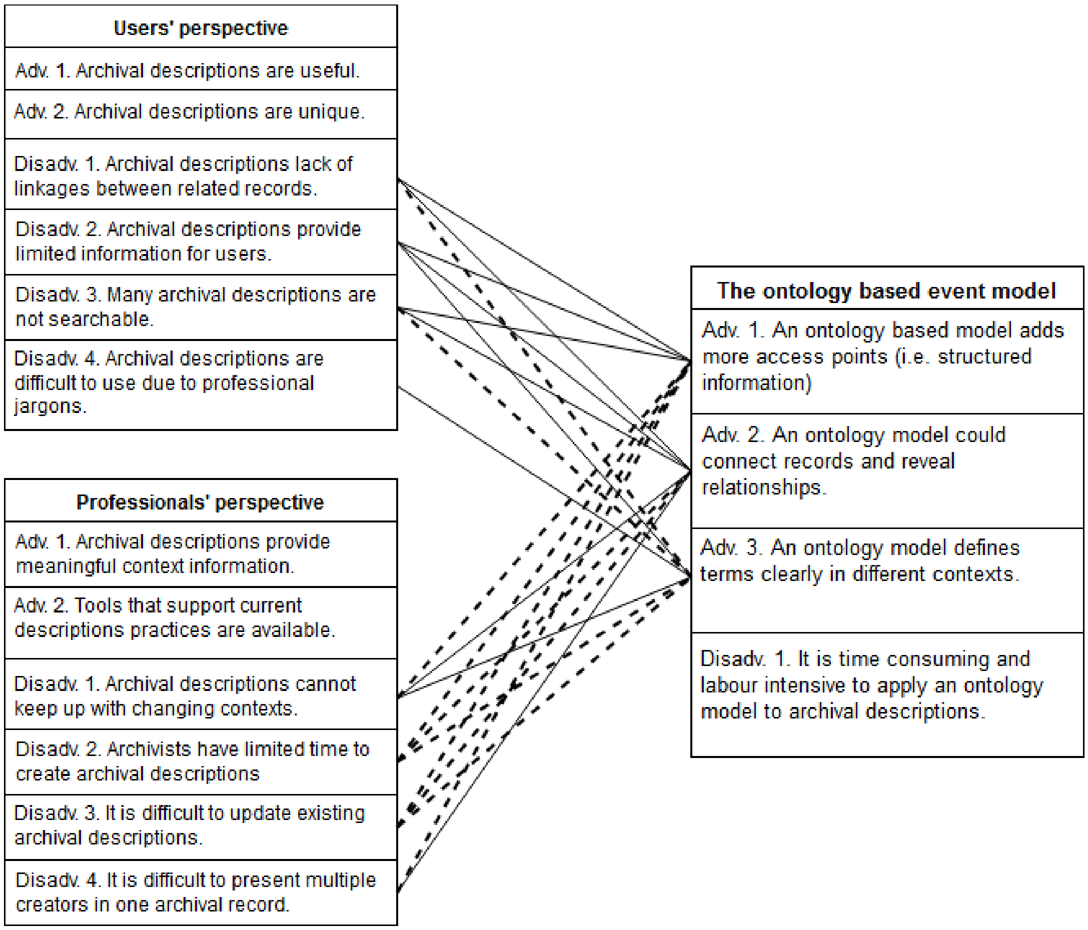


Figure 6.10 Mapping between the disadvantages of archival descriptions and advantages of the ontology model

Figure 6.10 shows the mappings between the disadvantages of archival descriptions and advantages of the ontology model. Ontology – Adv. 1 (the ontology based model adds more access points) can directly (denoted by solid lines) mitigate or solve the problem related to User – Disadv. 1 (Archival descriptions lack linkages between related records.) and User – Disadv. 2 (archival descriptions provide limited information for users). Since the use of an ontology-based event model implies that archival descriptions are represented as linked data, archival descriptions can become more searchable (User – Disadv. 3.) and be utilized to execute queries by using SPARQL (*SPARQL* Protocol and RDF Query Language). Thus, Ontology – Adv. 1 directly links to User –

Disadv. 3. Furthermore, Ontology – Adv. 1 may help update changing contexts in archival descriptions, solve limited time issues, help update existing archival descriptions, and better present multiple creators in a record. It may speed up the process of describing archival records depending on whether automation tools are well-designed. Thus, Ontology – Adv. 1 is indirectly (denoted by dotted lines) associated with Prof. – Disadv. 1 (Archival descriptions cannot keep up with changing contexts), Prof. – Disadv. 2 (Archivists have limited time to create archival descriptions), Prof. – Disadv. 3 (It is difficult to update existing archival descriptions), and Prof. – Disadv. 4 (It is difficult to present multiple creators in one archival record).

Ontology – Adv. 2 (An ontology model could connect records and reveal relationships) can directly facilitate linking related elements of records such as people, time, and place and aggregate them together. Thus, it has a direct impact on the User – Disadv. 1. (Archival descriptions lack linkages between related records) and can provide more information to users which address the User – Disadv. 2. (Archival descriptions provide limited information for users). It directly affects the search results of archival descriptions (User– Disadv. 3 (Many archival descriptions are not searchable)). Thus, it is directly related to User– Disadv. 1 – 3. Ontology – Adv. 2 links records and reveal relationships that can directly have an impact on Prof. – Disadv. 1 (Archival descriptions cannot keep up with changing contexts) and Prof. – Disadv. 4 (It is difficult to present multiple creators in one archival record). In addition, if an ontology model can connect records with more relationships, it may help with the limited time available to create archival descriptions (Prof. – Disadv. 2) and to better update existing archival descriptions (Prof. – Disadv. 3). Thus, Ontology – Adv. 2 indirectly links to Prof. – Disadv. 2 and 3.

Ontology – Adv. 3 (An ontology model defines terms clearly in different contexts) can directly provide more information to users, which helps provide more information for users (User

– Disadv. 2 , and mitigate or solve the disadvantages related to the use of professional jargon (User – Disadv. 4 (Archival descriptions are difficult to use due to professional jargon)). It also helps keep up with changing contexts (Prof. – Disadv. 1). Thus, Ontology – Adv. 3 are directly linked toProf. – Disadv. 1. In addition, Ontology – Adv. 3 is indirectly contributed to the lack of linkages (User 1 – Disadv. 1) and search issues (User – Disadv 3). It may also help with limited time (Prof. – Disadv. 2) and updating issues (Prof. – Disadv. 3). Thus, Ontology – Adv. 3 indirectly contributes to Prof. – Disadv. 2 and 3.

Through the two mapping tables, it is clear that the disadvantages of existing archival descriptions can be solved or at least mitigated with the advantages of the ontology model in one-to-one or one-to-many relationships. Recall that the participants of the two groups seem to perceive that the proposed ontology model is appropriate in fulfilling their needs. Thus, to make current archival descriptions more accessible, it is necessary, and archivists should explore a new way to better present archival descriptions. In addition, as the second theme of the focus group study, users have high expectations of the ontology model, which reflects the fact that ontology technologies may eliminate barriers for using archival descriptions, provide more access points to records and enhance search and retrieval results for end users.

Regarding the disadvantage of using the ontology model, there is a high learning curve for archivists to learn and apply ontology technologies. In addition, archivists are pressed for time and dealing with a large number of archival materials. They have limited time to create archival descriptions. Given the circumstances, it is important to have supporting tools available to implement the ontology model that can alleviate but not make the situation worse. If semi-automatic or automatic tools are implemented to facilitate the adoption of new ontology-based archival descriptions, aid the daily work of creating archival descriptions, and fix consistency problems and wrong mark-ups, archivists would be willing to implement the ontology model.

The findings of the focus group study demonstrate that users are open to technologies and models that can improve archival descriptions, especially those representing archival context. The implementation of ontology models within the existing archival descriptions requires technical and institutional supports including the construction of automatic tools, integration with existing archival information systems, and the cooperation of archivists and institutions. This is a practical challenge for adopting the proposed model as well as potential issues that need to be addressed in future research.

In summary, while the disadvantages identified by users reflect the problems in archival descriptions (especially archival context) when users utilize existing archival descriptions, the disadvantages identified by professionals reveal the problems related to professionals' creations of archival descriptions. The first mapping table shows that issues identified by users and professionals are interconnected. The second mapping table shows that the disadvantages that users and professionals identified can be mapped to each other and mitigated with the advantages of the ontology model.

6.4. Issues in ontology building procedures

In this study, methontology has been selected as a comprehensive method for building and implementing the event ontology. During the procedures for building a new ontology, two main concerns have been raised during implementation: 1) how to carry out the specification activity, and 2) how to select an upper ontology.

As explained in Chapter 4, section 4.2, Study 2 - Ontology Development and Representation of Archival Descriptions, methontology is the method used in Study 2. Concerning development activities, methontology consists of five phases including specification, conceptualization, formalization/integration, implementation, and maintenance (Fernández et al., 1997). Each phase has a respective outcome. Specification will identify a primary objective, purpose, and scope of the ontology. Among the list of outcomes, the ontology requirement specification is the outcome of the specification phase. Conceptualization is to construct the domain knowledge based on the outcomes of the specification phase. The glossary of terms, ad-hoc binary relation table, and concept dictionary are the outcome of the conceptualization phase. In the formalization phase, the informally perceived view of a domain is organized and converted into a formal or semi-computable model. Implementation transforms the conceptualization into an ontology language (e.g., Web Ontology Language). The event ontology for archival context (EOAC) is the outcome of the implementation phase.

Among the five phases, the ontology specification activity was the hard task in which knowledge acquisition occurs and requirements, purpose, and specifications of ontology are captured in the process of building an ontology. The specification process consists of seven tasks (i.e., to identify the purpose, intended users, intended uses, ontology requirements, to validate ontology requirements, to prioritize ontology requirements, and to extract terminology). In particular, in Tasks 3 (identify intended uses) and 4 (identify ontology requirements), the identification of intended uses and competency questions required describing the requirements of the ontology. In carrying out these tasks, there exist no detailed guidelines provided by the methontology. To address this issue, the use case diagrams of Unified Modelling Language (UML) was adopted because it is useful in describing "a set of use cases that systems (subject) perform in collaboration with one or more external users of the system (actors)" (Papajorgji, Pinet, Roussey, Brun, & Vigier, 2009; UML, 2009, para. 1). To figure out the competency questions, along with the case diagrams, a mind map tool was utilized. Using the mind map tool is advantageous for facilitating the convergence of concepts and removing distractions. These two tools help complement the specification activity and make the activity easier to carry out.

Secondly, in this study, an event-based ontology needs to be aligned with an upper ontology. This task is used to improve "semantic interoperability of ontologies" (Hoehndorf, 2010, para. 6). As methontology does not explicitly include how to select and use an upper ontology, the selection of an upper ontology was carefully taken to ensure interoperability. It also eliminates the vagueness of the domain ontology and situates the domain ontology upon a solid foundation. When selecting an upper ontology, the following features are usually considered: "ontological commitments (the philosophical choices taken), representation languages (the languages used to represent an ontology), software engineering properties, subject domain and applications" (Khan & Keet, 2012, p. 240). In addition, when assessing upper ontologies, four kinds of criteria are usually considered: licensing (e.g., proprietary or open), structure (e.g., flat or hierarchical), maturity (e.g., date of origin, development state, and level of adoption), and miscellaneous (Semy, Pulvermacher & Obrs, 2004).

Among these criteria, this study focused on structure, maturity, and ontological commitments. In terms of structure, the structure of the selected upper ontology should be flexible and extensible. The selected upper ontology should be mature, stable and well adopted in many domains and applications. The ontological commitments of the selected upper ontology focused on how objects, processes, time, and space are treated in the ontology. Based on the consideration of the criteria, the Basic Formal Ontology (BFO) was selected as the most appropriate one for the upper ontology of this study.

In summary, the findings of the research are presented to answer research questions. Archival context is multi-faceted and has multi-layered constructs. To examine the complexity of archival context, a three-level approach (i.e., micro-, meso- and macro-levels) is proposed as the extended findings of this research. At the micro-level, two views of archival context are presented, such as static and fluid views. The static view of archival context captures the core entities and relations of archival context. The fluid view of archival context represents the fluid nature of human activities. At the meso-level, three kinds of contexts are identified, including creation context, description context, and usage context. At the macro-level, CHAT has been utilized to identify entities, and a modified eight-step procedure is adopted based on Mwanza's model (Mwanza, 2002). In the focus group study, the responses of the two groups are discussed. The users and professionals identified the advantages and disadvantages of existing archival descriptions and the event-based ontology model, respectively. The two kinds of mappings are made to show that the raised issues of two groups correspond with each other. The two mappings indicate that the advantages of the ontology model can be mitigated to address the disadvantages of archival descriptions. Finally, the two issues related to ontology building procedures are discussed, such as knowledge acquisition in the specification stage and the selection of an upper ontology.

The next chapter presents the summary, contributions, limitations, and future directions of the research.

CHAPTER 7. CONCLUSION AND FUTURE DIRECTIONS

This research examines how archival contexts of archival descriptions are represented by using ontology technologies. Based on the findings of this research, a comprehensive conceptual model has been proposed and discussed. In this chapter, a summary of this research is presented in the first section. The contributions of the research are explained in the second section. Future research is then explained in the third section.

7.1. Summary of the research

This research set out to address the overarching question*: How can archival contexts be improved for searching, using and understanding in web environments*? Specifically, this research investigates the essential components of archival contexts and the way in which archival contexts can be represented by ontology technologies. A multi-method approach was adopted to answer the addressed research questions. First, a content analysis was used to identify essential contextual entities and relations of archival contexts in existing archival descriptions. Second, a methontology was adopted to build an event ontology for archival contexts (EOAC). Third, a focus group study elicited users' responses to the representation of archival contexts.

The findings of this research acknowledge that archival contexts are important. The findings further divided archival contexts into three kinds: creation context, description context, and usage context. This study identified essential entities and their relationships in the creation context and proposed static and fluid views of the creation context. More specifically, archival contextual entities identified in this study are *Agent, Artifact, Community, Event, Feature, Function, Place, Situation, Relation, Role, Rule, Time, and Tool.* Identified relations include the essential relationships between the contextual entities, such as kinship relations, social relations, and other agent-related relations.

The EOAC was developed through the stages of methontology: specification, conceptualization, implementation, formalization, and evaluation. The direct outcome of the methontology procedure (i.e. the EOAC) explicitly represents contextual entities and their relationships.

To assess the appropriateness of the ontology, focus group studies were conducted with professional archivists and user groups. The two groups identified the advantages and disadvantages of existing archival descriptions and the event-centred ontology model. Both groups identified the limitations of existing archival descriptions from professional and user perspectives and seemed to perceive that the proposed event ontology is appropriate for archival contexts.

The findings of the study provide a comprehensive conceptual model for identifying and representing archival contexts at the three levels (i.e., micro-, meso-, and macro levels). At the micro level, essential contextual entities and relations can be captured in the static and fluid views at the creation context. The static view is turned into an event ontology for archival context through the ontology building procedure (i.e., methontology). The static view can be served as a simplified conceptual model for archival contexts. At the meso level, the recognition of different contexts emphasizes the importance of the involvement of archivists. At the macro level, the recognition of higher level entities *Community, Tool*, and *Rule* includes the socio, cultural, technological aspects for archival contexts.

The EOAC enables the transformation of archival contexts into interconnected Resource Description Framework (RDF) statements, which change blocks of texts of archival contexts into interconnected forms. The EOAC improves linking within records across collections and enhances the findability of archival contexts within institutions as well as across institutional boundaries.

7.2. Contributions of the study

The findings of this study contribute to the advancement of knowledge at both the theoretical and methodological levels. This study also contributes to professional fields in archives. Theoretical, methodological, and professional aspects of the contributions are explained below.

7.2.1. Theoretical contributions

This research contributes to the recognition of the importance of archival contexts in the field of archival science. This research further emphasizes that archival contexts are the essential core of archival descriptions. This is important to shift from somewhat superficial descriptions of archival records to the constructive archival contexts of archival records. The EOAC is the core ontology for identifying archival contexts and provides the first comprehensive ontology, specifically for archival contexts. This ontology is a starting point to represent archival contexts ontologically in a Linked data approach. The use of ontology technologies can be further expanded to use other Semantic Web technologies and semantic annotations. The application of ontology technologies in the Semantic Web provides an innovated way to systematically analyze archival contexts and brings about new ways to enhance both fields of knowledge representation and archival science.

This research has proposed an event-oriented approach, which is different from the existing object-centred approach that still presents challenges in dealing with various contextual entities (e.g., biographical information and administrative histories) when describing archival records. The event-centred approach has the advantage of connecting events, situations, and other contextual entities. The proposed event-oriented approach not only utlizes object-centred modelling, but also promotes events as equally important as objects. The shift from an object-centred to event-centred approach will have significant impacts on archival description theory, which is a starting point from which a new metadata framework could be further developed, as Niu (2015) has advocated. The EOAC ontology is the first event-centred model explicitly representing archival contexts and has great potential to be further developed to improve the linkages between archival descriptions, foster new applications based on events, and enhance the search functionality of archival information systems.

This research contributes to the construction of a comprehensive conceptual model for understanding archival contexts at three levels. At the macro level, Cultural-Historical Activity Theory (CHAT) was introduced to capture social, cultural, and technological contexts in archival descriptions. The development of the modified eight-step model could help identify and capture the broad contexts. The model can be further developed into a new appraisal approach that can enhance and advance appraisal theory in archives.

At the meso level, the typology of contexts (i.e., creation, description, and usage contexts) is identified to explicitly categorize contexts that makes it clear how to identify the possible interactions between contexts and stakeholders in each context. The categorization makes it easier for archivists or developers to identify the boundaries of each context and advance the understanding of archival contexts.

At the micro level, the creation context extends an understanding of how to capture the record creation process in the two types of views (the static and fluid views). In this way, the dynamic of creation context can be captured through snapshots (i.e., static views) and spans (i.e., fluid views). Moreover, this study made the Event entity important and recognized the fluid nature of history, process, and activities, which is an important shift that makes archival descriptions more structured, explicit and intuitive.

This research provides not only a model of comprehensive archival contexts but an ontology that can transform archival contexts into Linked Data. Moreover, this research introduced an innovative event-centred approach and emphasized events as being as important as objects. This shift has significant implication on how archival context can be organized and managed differently. Finally, the adoption of the Cultural-Historical Activity Theory (CHAT) helps model cultural, social, and technological aspects of contexts when examining contexts. CHAT has been often used in the fields of human-computer interaction and social studies. This is the first time that CHAT has been used to capture broader contexts in archival descriptions at a higher level. All of these original contributions demonstrate how knowledge representation, an event-centred approach and CHAT can provide innovative ways to model archival contexts.

7.2.2. Methodological contributions

The originality of the research is also in its methodology. First, the use of a mixed-method approach to examine archival contexts sheds light on how archival contexts and archival descriptions can be organized and represented in a clear and explicit way.

Second, this is the first time that methontology has been applied to build an event ontology for archival contexts in the field of archives. Methontology has been modified and expanded for this research. This study used the UML (Unified Model Language) for identifying the specifications of the ontology in a modified procedure. Moreover, in the knowledge acquisition process, knowledge organization systems such as control vocabulary, thesaurus, and a glossary of terms have been introduced and utilized to broaden the sources of knowledge acquisition. This approach takes advantage of methontology and further enhances and improves the validity and reliability of the method. The use of a focus group study is also meaningful in evaluating the EOAC, as well as to better meet users' expectations.

Third, the use of ontology can transform archival contexts, traditionally locked in large blocks of texts, into RDF statements with the ability to connect with archival records and beyond.

With all of these contributions, the understanding of archival contexts, descriptions, and knowledge acquisitions in building an ontology have been deepened. This research presents a new way to understand archival contexts and transform archival contexts into Linked Open Data that can significantly improve the findability of archival records.

7.2.3. Contributions to professional fields

This study proposed a comprehensive archival contexts model and developed the EOAC ontology, which has significant practical implications for professionals, users, system designers, and policymakers.

For professionals, the findings contribute to a better understanding of archival contexts, providing a comprehensive view of archival contexts to professionals using archives (e.g., archivists, records professionals, and historians). The three levels of archival contexts (i.e., micro-, meso-, and macro levels) make archival contexts explicit when they handle archival records. The eight-step model at the macro level helps them clearly facilitate an appraisal of archival records in their work.

The outcome of the EOAC provides a basic ontology platform for professionals when implementing ontology technologies to archival descriptions. This can transform archival descriptions into structured RDFs by presenting events and their interlinking relationships within records and beyond institutional boundaries.

For users, the findings of this research can highly improve the user experience of searching and using archival descriptions. The EOAC transforms archival descriptions into structured forms with more access points that have the potential to improve search functions and results. Thus, the linked records in the event model are more accessible to users.

For system designers, the findings of this research can offer a categorization of archival contexts through the EOAC model, which helps them design archival contexts to be more comprehensive and easier to understand and helps integrate them into new archival information systems. The EOAC model, along with open linked data, provides a framework for linking archival descrptions within and beyond the archival insitutional boundaries. The model also has the advantage of clearly and systematically describing and managing metadata in activities/processes of individuals, families, and cooperate-bodies compared to object-centred approaches.

For policymakers, the findings of this research provide a practical way to transform archival descriptions to build linkages within and beyond archival records. The adoption of the EOAC in archives can help archival communities adopt Linked Open Data and move towards a union catalogued database of archival descriptions.

7.3. Future research directions

Through the examination of archival contexts in this research, there are additional issues related to archival context. This section explains selected issues on which further investigation will be needed.

First, to practically use the EOAC model, implementation-related issues are important. As the event-ontology model was perceived in the focus groups to be appropriate, but with certain implementation conditions, implementation is the next step to transform existing archival descriptions in a practical setting. Although the complete aspects of implementation are not within the scope of this study, some further suggestions would be helpful for implementation.

To apply the EOAC ontology to archival descriptions, semantic annotation could be used. Semantic annotation refers to "[a]n annotation assign[ed] to an entity, which is in the text, a link to its semantic description. Semantic annotation is referent to an ontology" (Talantikite, Aissani, & Boudjlida, 2009, p. 1109). Specifically, it is one type of application to "link ontologies to the original information sources" (Lin, 2008, p. 23).

Figure 7.1 shows that text snippets are annotated with different annotations (i.e., Features, ArchivalFonds, Donation, Organization, Person, and Interval).

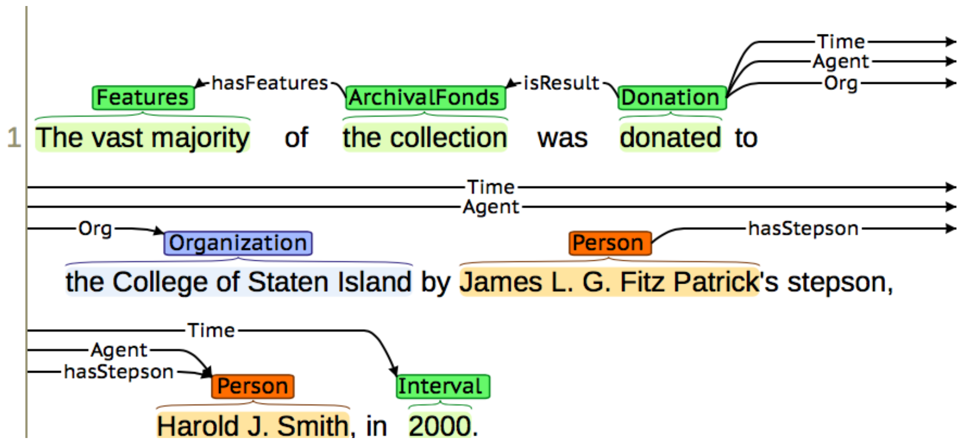


Figure 7.1 Semantic annotations

The annotations are associated with text snippets. Relations, such as hasFeatures, isResult, and hasStepson, further connect related things together. For example, Harold J. Smith is not just textual information, but refers to an instance of a person and a stepson of another person. Semantic annotations help establish these connections. These connections are important parts in the foundation of a semantic search engine so that ontologies will link to the sources. This enables the building of smart search engines, rather than pure text-based search engines.

Semantic annotations powered by archival information systems can be built on certain open source archival information systems such as AtoM or ArchivalSpace. The EOAC and semantic annotation can be built as add-on modules of archival information systems. With such systems, two types of studies can be carried out to evaluate: 1) how effectively professionals use semantic annotations with the EOAC; 2) how effective it works from the users' perspectives.

With the new development of machine learning and natural language processing, the semantic annotation process (i.e., the creation process of applying the EOAC model to archival descriptions) could be semi-automatically or automatically accomplished. The process can be divided into smaller tasks, such as named-entities recognition, event detection, event extraction, text summarization, and information extraction. To accomplish the tasks, there are at least three types of approaches: pattern-based systems, feature-based (machine learning based on engineered features) and neural-based approaches (Boroş, 2018). The feature-based and neural-based often require high quality training sets in order to train models and yield meaningful results, although such sets are very costly to build. Big data provides opportunities to train models without massive labeled training sets, such as Stanford University's DAWN project (Stanford University, 2018). It is hoped that the plugins (add-on modules) could be further extended by using machine learning and natural language processing technologies.

Second, examining the archival contexts of digital archives could be considered. In this work, traditional archives in paper format would be treated. With more and more born-digital materials being created, archives need to shift to digital archives. Born digital archival records are created in an electronic format that inherits the characteristics of the format. The archival context of digital archives needs to reflect the characteristics of digital archives.

PROV is a W3C recommendation to "express provenance records, which contain descriptions of the entities and activities involved in producing and delivering or otherwise influencing a given object" (Gil & Miles, 2013, para. 1). The term "provenance" is defined as "a record that describes the people, institutions, entities, and activities involved in producing, influencing, or delivering a piece of data or a thing" (Moreau & Missier, 2013, para. 1) and it is different from the definitions of provenance in archives. PROV and its Model (OPM) are more "focusing on the provenance history" (especially technological contexts) of a digital object (Bountouri, 2017, p. 46). While the EOAC model covers broader contexts, the combined use of EOAC and PROV further extends the two models to meet broader needs. How these may be combined and the alignment of EOAC with PROV could be explored through case studies in which PROV metadata and their relationships with EOAC, archival context, and archival descriptions are examined.

Lastly, validation is a vital step in building an ontology, as explained in Chapter 4 section 4.2.5, Stage 5. maintenance. Archival user studies have focused on archival users who investigate how they search, locate, and use archives. Since the EOAC can fundamentally shape the way archival contexts are presented and retrieved and can improve information retrieval, archival user studies were adopted to evaluate the applications of EOAC in this study. For further studies, archival user studies could be carried out to test whether different types of professionals (e.g., archivists, librarians, historians, genealogists, etc.) respond differently, and which parts of the EOAC could be improved. In addition, the appropriateness of the EOAC application could be tested with multiple user groups searching and retrieving archival descriptions and then evaluating the users' performances (e.g., time, precision, recall) when they conduct search tasks against archival descriptions.

REFERENCES

Abraham, T. (2015). Repositories of primary sources. Retrieved March 28, 2018, from https://web.archive.org/web/20150509061807/http://webpages.uidaho.edu:80/special-collections/other.repositories.html

Adams, M. O. (2007). Analyzing archives and finding facts: use and users of digital data records. *Archival Science*, *7*(1), 21–36. https://doi.org/10.1007/s10502-007-9056-4

Agarwal, N., Ford, K. H., & Shneider, M. A. (2005). Sentence boundary detection using a maxent classifier.

Al-Baltah, I. A., Ghani, A., RAHMAN, W., & Atan, R. (2014). A comparative study on ontology development methodologies towards building semantic conflicts detection ontology for heterogeneous web services. *Research Journal of Applied Sciences, Engineering and Technology*, *7*(13), 2674–2679.

Allen, B., & Reser, D. (1990). Content analysis in library and information science research. *Library & Information Science Research*, *12*(3), 251–260.

Allen, J. F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM*, *26*(11), 832–843.

Anderson, I. G. (2004). Are you being served? historians and the search for primary sources. *Archivaria*, *58*, 81–129.

ArchivesSpace. (2018). ArchivesSpace. Retrieved March 26, 2018, from http://archivesspace.org/

Archivists' Toolkit project. (2009). Archivists' Toolkit | For archivists by archivists. Retrieved March 26, 2018, from http://www.archiviststoolkit.org/

Arp, R., Smith, B., & Spear, A. D. (2015). *Building ontologies with basic formal ontology*. Cambridge Mass.: The MIT Press.

Artefactual Systems Inc. (2015). AtoM: Open Source Archival Description Software. Retrieved March 26, 2018, from https://www.accesstomemory.org/en/

Baader, F., Calvanese, D., Mcguinness, D. L., Nardi, D., & Patel-Schneider, P. F. (2007). *The description logic handbook* (2nd ed.). New York, NY: Cambridge University Press.

Bahde, A. (2017). Conceptual data visualization in archival finding aids: preliminary user responses. *Portal: Libraries and the Academy*, *17*(3), 485–506. https://doi.org/10.1353/pla.2017.0031

Basic Formal Ontology. (2011). Retrieved from http://www.ifomis.org/bfo

Bastian, J. A. (2003). *Owning memory: how a caribbean community lost its archives and found its history*. Westport, CT: Libraries Unlimited.

Bastian, J. A., & Alexander, B. (Eds.). (2009). *Community archives: the shaping of memory*. London, England: Facet Publishing.

Bearman, D. (1995). Archival strategies. *American Archivist*, *58*(4), 380–413.

Bechhofer, S., Harmelen, F. van, Hendler, J., Horrocks, I., McGuinness, D. L., Patel-Schneider, P. F., & Stein, L. A. (2004). *Owl web ontology language reference*. World Wide Web Consortium. Retrieved from https://www.w3.org/TR/owl-ref/

Bennett, B., & Galton, A. (2004). A unifying semantics for time and events. *Artificial Intelligence*, *153*(1–2), 13–48. https://doi.org/10.1016/j.artint.2003.02.001

Benson, A. C. (2009). The archival photograph and its meaning: formalisms for modeling images. *Journal of Archival Organization*, *7*(4), 148–187. https://doi.org/10.1080/15332740903554770

Berkhofer, R. F. (1995). *Beyond the great story: history as text and discourse*. Cambridge, MA: Belknap Press of Harvard University Press.

Berners-Lee, T. (2000, December). *Semantic web - xml2000*. Retrieved from http://www.w3.org/2000/Talks/1206-xml2k-tbl/Overview.html

Berners-Lee, T. (2006). Linked data-design issues. Retrieved from http://www.w3.org/DesignIssues/LinkedData.html

Berners-Lee, T., Hendler, J. A., & Lassila, O. (2001, May). The semantic web. *Scientific American*, *284*(5), 34–43.

Bizer, C., Heath, T., Idehen, K., & Berners-Lee, T. (2008). Linked data on the web. In *Proceedings of the 17th International Conference on World Wide Web* (pp. 1265–1266). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1367760

Boettger, R. K., & Palmer, L. A. (2010). Quantitative content analysis: its use in technical communication. *IEEE Transactions on Professional Communication*, *53*(4), 346–357. https://doi.org/10.1109/TPC.2010.2077450

Borst, W. N. (1997). *Construction of engineering ontologies for knowledge sharing and reuse*. University of Twente, Enschede, Netherlands.

Bountouri, L. (2017). *Archives in the digital age: standards, policies and tools* (1st ed.). London, England: Chandos Publishing.

Brachman, R., & Levesque, H. (2004). *Knowledge representation and reasoning*. San Francisco, CA: Morgan Kaufmann.

Bradley, J. (1993). Methodological issues and practices in qualitative research. *The Library Quarterly: Information, Community, Policy*, *63*(4), 431–449.

Breitman, K. K., Casanova, M. A., & Truszkowski, W. (2007). *Semantic web concepts, technologies and applications*. New York; London: Springer. Retrieved from http://dx.doi.org/10.1007/978-1-84628-710-7

Brickley, D. (2004). *W3c semantic web interest group: basic geo (wgs84 lat/long) vocabulary*. Retrieved from http://www.w3.org/2003/01/geo/

Brickley, D., & Guha, R. (2004). *RDF vocabulary description language 1.0: rdf schema*. Retrieved from http://www.w3.org/TR/rdf-schema

Brickley, D., & Miller, L. (2014). FOAF vocabulary specification 0.99. Retrieved from http://xmlns.com/foaf/spec/

Brusa, G., Laura Caliusco, M., & Chiotti, O. (2008). Towards ontological engineering: a process for building a domain ontology from scratch in public administration: towards ontological engineering. *Expert Systems*, *25*(5), 484–503. https://doi.org/10.1111/j.1468-0394.2008.00471.x

Bureau of Canadian Archivists, Planning Committee on Descriptive Standards. (2008). *Rules for archival description*. Ottawa, Canada: Bureau of Canadian Archivists.

Cardoso, J. (2007). The semantic web vision: where are we? *IEEE Intelligent Systems*, *22*(5), 84–88. http://doi.ieeecomputersociety.org/10.1109/MIS.2007.97

Chapman, J. C. (2010). Observing users: an empirical analysis of user interaction with online finding aids. *Journal of Archival Organization*, *8*(1), 4–30. https://doi.org/10.1080/15332748.2010.484361

Charmaz, K. (1995). Grounded theory. In J. A. Smith, R. Harré, & L. V. Langenhove (Eds.), Rethinking Methods in Psychology. London, England: SAGE Publications Ltd. https://doi.org/10.4135/9781446221792

Casati, R., & Varzi, A. (2015). Events. In E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy (Winter 2015). Retrieved from https://plato.stanford.edu/archives/win2015/entries/events/

Chaudron, G. (2008). The potential of "function" as an archival descriptor. *Journal of Archival Organization*, *6*(4), 269–287. https://doi.org/10.1080/15332740802533313

CIDOC CRM Special Interest Group Working Group. (2011). *Cidoc crm*. Retrieved from http://www.cidoc-crm.org/

Clinton Digital Library. (2018). Browse items · Clinton Digital Library. Retrieved June 20, 2018, from https://clinton.presidentiallibraries.us/items/browse?collection=82&sort_field=Dublin+Core%2CTitle&sort_dir=a

Coats, L. R. (2004). Users of ead finding aids: who are they and are they satisfied? *Journal of Archival Organization*, *2*(3), 25–39. https://doi.org/10.1300/J201v02n03_03

Cole, M. (1996). *Cultural psychology: a once and future discipline*. Cambridge, MA: Belknap Press of Harvard University Press.

Comisión de Normas Españolas de Descripción Archivística. (2012). *Report on the work of cneda (2007-2012): toward a conceptual model for archival description in spain*. Spain. Retrieved from https://www.mecd.gob.es/cultura-mecd/dms/mecd/cultura-mecd/areas-cultura/archivos/mc/cneda/documentacion/ReportCNEDA_11_07_2012-pdf/ReportCNEDA_11_07_2012.pdf

Conway, P. (1986). Facts and frameworks: an approach to studying the users of archives. *The American Archivist*, *49*(4), 393–407.

Cook, M. (1999). *Managing archives.* London: International Records Management Trust and International Council on Archives. Retrieved from http://irmt.org/documents/educ_training/public_sector_rec/IRMT_manage_archives.pdf

Cook, T. (1997). What is past is prologue: a history of archival ideas since 1898, and the future paradigm shift. *Archivaria*, *43*, 17–63.

Cook, T. (2001). Archival science and postmodernism: new formulations for old concepts. *Archival Science*, *1*, 3–24.

Corcho, O., Fernández-López, M., & Gómez-Pérez, A. (2003). Methodologies, tools and languages for building ontologies. where is their meeting point? *Data & Knowledge Engineering*, *46*(1), 41–64. http://dx.doi.org/10.1016/S0169-023X(02)00195-7

Corcho, O., Fernández-López, M., & Gómez-Pérez, A. (2006). Ontological engineering: principles, methods, tools and languages. In *Ontologies for software engineering and software technology* (pp. 1–48). Berlin, Heidelberg, Gemany: Springer Berlin Heidelberg.

Corcho, O., Fernández-López, M., Gómez-Pérez, A., & López-Cima, A. (2005). Building legal ontologies with methontology and webode. In *Law and the semantic web: legal ontologies, methodologies, legal information retrieval, and applications* (pp. 142–157). Berlin, Heidelberg, Gemany: Springer.

Cox, R. J. (1994). An analysis of archival research, 1970-92, and the role and function of the American archivist. *The American Archivist*, *57*(2), 278–288. https://doi.org/10.17723/aarc.57.2.533tt46jh6741755

Cox, R. J. (2007). Revisiting the archival finding aid. *Journal of Archival Organization*, *5*(4), 5–32. https://doi.org/10.1080/15332740802153245

Coyle, K. (2010). Understanding the semantic web: bibliographic data and metadata. *Library Technology Reports*, *46*(1), 1–34.

Creswell, J. W. (2003). *Research design: qualitative, qunatitative, and mixed methods approaches* (2nd ed.). Thousand Oaks, CA, USA: Sage Publications.

Cristani, M., & Cuel, R. (2005). A survey on ontology creation methodologies: *International Journal on Semantic Web and Information Systems*, *1*(2), 49–69. https://doi.org/10.4018/jswis.2005040103

Crotty, M. (1998). *The foundations of social science research: meaning and perspective in the research process*. London, UK: SAGE Publications.

Dahab, M., Hassan, H., & Rafea, A. (2008). Textontoex: automatic ontology construction from natural English text. *Expert Systems with Applications*, *34*(2), 1474–1480. https://doi.org/10.1016/j.eswa.2007.01.043

Daines, J. G., & Nimer, C. L. (2011). Re-imagining archival display: creating user-friendly finding aids. *Journal of Archival Organization*, *9*(1), 4–31. https://doi.org/10.1080/15332748.2011.574019

Daniels, M., & Yakel, E. (2010). Seek and you may find: successful search in online finding aid systems. *The American Archivist*, *73*(2), 535–568. https://doi.org/10.17723/aarc.73.2.p578900680650357

Daniels, M., & Yakel, E. (2013). Uncovering impact: the influence of archives on student learning. *The Journal of Academic Librarianship*, *39*(5), 414–422. https://doi.org/10.1016/j.acalib.2013.03.017

Davies, D. (2003). *Thinking outside the box: re-imaging archival description with "series" system* (Master). University of Manitoba, Winnipeg, Canada.

Davis, A. M. (1993). *Software requirements: objects, functions, and states*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Davis, I. (2010). Relationship: a vocabulary for describing relationships between people. Retrieved October 10, 2018, from http://vocab.org/relationship/

Davis, R., Shrobe, H., & Szolovits, P. (1993). What is a knowledge representation. *AI Magazine*, *14*(1), 17–33.

Day-Richter, J., Harris, M. A., Haendel, M., & Lewis, S. (2007). OBO-Edit—an ontology editor for biologists. *Bioinformatics*, *23*(16), 2198–2200. https://doi.org/10.1093/bioinformatics/btm112

Derczynski, L. R. A. (2017). *Automatically Ordering Events and Times in Text*. Springer International Publishing. Retrieved from https://www.springer.com/gp/book/9783319472409

Pan-Canadian Documentary Heritage Network (PCDHN). (2018). *Out of the Trenches: A Linked Open Data Project* (Version 2). University of Alberta Libraries Dataverse. https://doi.org/10.7939/DVN/URXSGC

Doerr, M. (2003). The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata. *AI Magazine*, *24*(3), 75–92.

Doerr, M., & Crofts, N. (1999). Electronic Esperanto: the role of the object oriented CIDOC reference model. In D. Bearman & J. Trant (Eds.), *Cultural Heritage Informatics 1999: Selected papers from ichim99* (pp. 157–173). Toronto, Canada: Archives & Museum Informatics.

Dombeu, J. V. F., & Huisman, M. (2011). Combining ontology development methodologies and semantic web platforms for e-government domain ontology development. *International Journal of Web & Semantic Technology (IJWesT)*, *2*(2), 12–25.

Dourish, P. (2004). What we talk about when we talk about context. *Personal Ubiquitous Computing*, *8*(1), 19–30. https://doi.org/10.1007/s00779-003-0253-8

Dow, E. H. (2009). Encoded Archival Description as a Halfway Technology. *Journal of Archival Organization*, *7*(3), 108–115.

Dublin Core Metadata Initiative. (2012). DCMI: DCMI metadata terms. Retrieved October 10, 2018, from http://dublincore.org/documents/dcmi-terms/

Duchein, M. (1983). Theoretical principles and practical problems of respect des fonds in archival science. *Archivaria*, *16*, 64–82.

Duff, W. M., & Harris, V. (2002). Stories and names: archival description as narrating records and constructing meanings. *Archival Science*, *2*(3), 263–285. https://doi.org/10.1007/BF02435625

Duff, W. M., & Stoyanova, P. (1998). Transforming the crazy quilt: archival displays from a users' point of view. *Archivaria*, *45*, 44–79.

Duff, W. M., Craig, B., & Cherry, J. (2004). Finding and using archival resources: a cross-canada survey of historians studying canadian history. *Archivaria*, *58*, 51–80.

Duranti, L. (1993). Origin and development of the concept of archival description. *Archivaria*, *35*, 47–54.

Ehrlinger, L., & Wöß, W. (2016). Towards a Definition of Knowledge Graphs. In M. Martin, M. Cuquet, & E. Folmer (Eds.), *Joint Proceedings of the Posters and Demos Track of the 12th International Conference on Semantic Systems - SEMANTiCS2016 and the 1st International Workshop on Semantic Change & Evolving Semantics (SuCCESS'16)* (Vol. 1695). Leipzig, Germany: CEUR-WS. Retrieved from http://ceur-ws.org/Vol-1695/paper4.pdf

Engel, R. J., & Schutt, R. K. (2012). *The practice of research in social work*. Thousand Oaks, CA: SAGE Publications.

Engeström, Y. (1987). *Learning by expanding: an activity-theoretical apporach to development research*. Helsinki, Finland: Orienta-Konsultit.

Engeström, Y. (1990). When is a tool? multiple meanings of artifacts in human activity. In *Learning, Working and Imaging: Twelve Studies in Activity Theory* (pp. 171–195). Helsinki, Finland: Orienta-Konsultit.

Engeström, Y. (1999). Communication, discourse and activity. *The Communication Review*, *3*(1–2), 165–185. https://doi.org/10.1080/10714429909368577

Europeana Project. (2013). *Europeana data model primer*. Retrieved from https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf

Evans, F. B., Harrison, D. F., Thompson, E. A., & Rofes, W. L. (1974). A basic glossary for archivists, manuscript curators, and records managers. *The American Archivist*, *37*(3), 415–433.

Evans, J., McKemmish, S., & Reed, B. (2010). Archival arrangement and description. In M. J. Bates & M. N. Maack (Eds.), *Encyclopedia of library and information sciences* (3rd ed., pp. 130–140). Boca Raton, FL: CRC Press.

Experts Group on Archival Description. (2016). *Records in contexts - a conceptual model for archival description*. International Council on Archives. Retrieved from https://www.ica.org/en/egad-ric-conceptual-model

Fakhroutdinov, K. (2009). UML Use Case Diagrams. Retrieved September 27, 2018, from https://www.uml-diagrams.org/use-case-diagrams.html

Fernández-López, M., & Gómez-Pérez, A. (2002). Overview and analysis of methodologies for building ontologies. *The Knowledge Engineering Review*, *17*(2), 129–156. https://doi.org/10.1017/S0269888902000462

Kaenampornpan, M., & O'Neill, E. (2004). Modelling Context: An Activity Theory Approach. In P. Markopoulos, B. Eggen, E. Aarts, & J. L. Crowley (Eds.), Ambient Intelligence (pp. 367–374). Berlin, Heidelberg: Springer Berlin Heidelberg.

Kohlbacher, F. (2006). The Use of Qualitative Content Analysis in Case Study Research. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *7*(1). https://doi.org/10.17169/fqs-7.1.75

Yu, C., & Luo, Y. (2012). Comparative Research on Methodologies for Domain Ontology Development. In D.-S. Huang, Y. Gan, P. Gupta, & M. M. Gromiha (Eds.), *Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence*

(Vol. 6839, pp. 349–356). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-25944-9_45

IEEE. (1996). *IEEE Standard for Developing Software Life Cycle Processes* (IEEE Std 1074-1995). https://doi.org/10.1109/IEEESTD.1996.79663

Morgan, D. L. (2006). Focus group. In V. Jupp (Ed.), *The Sage dictionary of social research methods* (pp. 121–123). Thousand Oaks, CA: Sage

Fernández-López, M., Gómez-Pérez, A., & Juristo, N. (1997). Methontology: from ontological art towards ontological engineering. In A. Farquhar & M. Gruninger (Eds.), *Proceedings of the Ontological Engineering AAAI-97 Spring Symposium Series* (pp. 33–40). Menlo Park, CA: AAAI Press.

Foot, K. A. (2014). Cultural-Historical Activity Theory: Exploring a Theory to Inform Practice and Research. *Journal of Human Behavior in the Social Environment*, *24*(3), 329–347. https://doi.org/10.1080/10911359.2013.831011

Foscarini, F. (2010). Understanding the context of records creation and use: 'hard' versus 'soft' approaches to records management. *Archival Science*, *10*(4), 389–407. https://doi.org/10.1007/s10502-010-9132-z

Foscarini, F. (2013). Record as social action: understanding organizational records through the lens of genre theory. *Information Research*, *18*(3). Retrieved from http://www.informationr.net/ir/18-3/colis/paperC08.html

Freund, L., & Toms, E. G. (2016). Interacting with archival finding aids. *Journal of the Association for Information Science & Technology*, *67*(4), 994–1008.

Furner, J. (2016). "Records in context" in context: a brief history of data modelling for archival description. In F. Foscarini, H. MacNeil, B. Mak, & G. Oliver (Eds.), *Engaging with Records and Archives: Histories and Theories* (pp. 41–62). London, UK: Facet Publishing.

Galton, A. (1995). Time and change for AI. In D. M. Gabbay, C. J. Hogger, & J. A. Robinson (Eds.), *Handbook of logic in artificial intelligence and logic programming (Vol. 4): epistemic and temporal reasoning* (pp. 175–240). Oxford, UK: Oxford University Press.

Galton, A. (2008). Experience and history: processes and their relation to events. *Journal of Logic and Computation*, *18*(3), 323–340. https://doi.org/10.1093/logcom

Galton, A., & Mizoguchi, R. (2009). The water falls but the waterfall does not fall: new perspectives on objects, processes and events. *Applied Ontology*, *4*(2), 71–107. https://doi.org/10.3233/AO-2009-0067

Gangemi, A., & Mika, P. (2003). Understanding the Semantic Web through Descriptions and Situations. In R. Meersman, Z. Tari, & D. C. Schmidt (Eds.), *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE* (pp. 689–706). Berlin, Heidelberg, Gemany: Springer Berlin Heidelberg.

Gašević, D., Djurić, D., & Devedžić, V. (2006). *Model driven architecture and ontology development*. Heidelberg, Gemany: Springer-Verlag.

Gerhold, D. (2013). Record offices online catalogues: an appraisal. *The Local Historian*, *43*(1), 57–63.

Gil, Y., & Miles, S. (2013). Prov model primer. Retrieved September 26, 2018, from https://www.w3.org/TR/prov-primer/

Gill, T. (2004). Building semantic bridges between museums, libraries and archives: the CIDOC conceptual reference model. *First Monday*, *9*(5). https://firstmonday.org/ojs/index.php/fm/article/view/1145/1065

Gilliland, A. J. (2012). Contemplating co-creator rights in archival description. *Knowledge Organization*, *39*(5), 340–346.

Gilliland-Swetland, A. J. (1998). Evaluation design for large-scale, collaborative online archives: interim report of the online archive of California evaluation project. *Archives and Museum Informatics*, *12*(3–4), 177–203.

Gómez-Pérez, A., Fernández-López, M., & Corcho, O. (2004). *Ontological engineering: with examples from the areas of knowledge management, e-commerce and the semantic web*. London ; New York: Springer.

Gracy, K. F. (2015). Archival description and linked data: a preliminary study of opportunities and implementation challenges. *Archival Science*, *15*(3), 239–294. https://doi.org/10.1007/s10502-014-9216-2

Green, H. E., & Lampron, P. (2017). User engagement with digital archives for research and teaching: a case study of emblematica online. *Portal: Libraries and the Academy*, *17*(4), 759–775. https://doi.org/10.1353/pla.2017.0045

Greenberg, S. (2001). Context as a dynamic construct. *Human-Computer Interaction*, *16*, 257–268.

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, *11*(3), 255–274. https://doi.org/10.3102/01623737011003255

Greene, M., & Meissner, D. (2005). More product, less process: revamping traditional archival processing. *The American Archivist*, *68*(2), 208–263.

Grenon, P., & Smith, B. (2004). Snap and span: towards dynamic spatial ontology. *Spatial Cognition & Computation*, *4*(1), 69–104.

Groth, P., & Moreau, L. (2013). Prov-overview. Retrieved May 2, 2018, from https://www.w3.org/TR/2013/NOTE-prov-overview-20130430/

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, *5*(2), 199–220.

Grüninger, M., & Fox, M. S. (1995). Methodology for the design and evaluation of ontologies. In *Workshop on Basic Ontological Issues in Knowledge Sharing*. Montreal. Retrieved from http://stl.mie.utoronto.ca/publications/gruninger-ijcai95.pdf

Guarino, N. (1998). Formal ontology and information systems. In N. Guarino (Ed.), *Proceedings of the 1st International Conference on Formal Ontologies in Information Systems (FOIS'98)* (pp. 3–15). Trento, Italy: IOS Press.

Gueguen, G., da Fonseca, V., Pitti, D., & Grimoüard, C. (2013). Toward an international conceptual model for archival description: a preliminary report from the international council on archives' experts group on archival description. *The American Archivist*, *76*(2), 567–584.

Guinebert, M., Yessad, A., Muratet, M., & Luengo, V. (2017). An ontology for describing scenarios of multi-players learning games: toward an automatic detection of group interactions. In É. Lavoué, H. Drachsler, K. Verbert, J. Broisin, & M. Pérez-Sanagustín (Eds.), *Data Driven Approaches in Digital Education* (pp. 410–415). Cham: Springer International Publishing.

Ham, F. (1984). Archival choices: managing the historical record in an age of abundance. *The American Archivist*, *47*(1), 11–22. https://doi.org/10.17723/aarc.47.1.v382727652114521

Hayes, P. (2004). RDF Semantics. Retrieved from http://www.w3.org/TR/2004/REC-rdf-mt-20040210/

Hickerson, H. (2001). Ten challenges for the archival profession. *The American Archivist*, *64*(1), 6–16. https://doi.org/10.17723/aarc.64.1.0rg6525nu1437v31

Hitzler, P., Krötzsch, M., & Sebastian, R. (2009). *Foundations of semantic web technologies*. London, England: Chapman & Hall/CRC.

Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P. F., & Rudolph, S. (2012). *OWL 2 web ontology language primer (second edition)*. World Wide Web Consortium. Retrieved from https://www.w3.org/TR/2012/REC-owl2-primer-20121211/

Hobbs, J. R., & Pan, F. (2006). *Time ontology in owl*. World Wide Web Consortium. Retrieved from https://www.w3.org/TR/owl-time/

Hoehndorf, R. (2010). What is an upper level ontology? Retrieved September 27, 2018, from http://ontogenesis.knowledgeblog.org/740

Holloway, I., & Wheeler, S. (2010). *Qualitative research in nursing and healthcare* (3rd ed.). Chichester, West Sussex, United Kingdom: Wiley-Blackwell.

Horridge, M. (2011). *A Practical Guide To Building OWL Ontologies Using Prot´eg´e 4 and CO-ODE Tools Edition 1.3* (p. 108). University of Manchester. Retrieved from http://mowl-power.cs.man.ac.uk/protegeowltutorial/resources/ProtegeOWLTutorialP4_v1_3.pdf

Hou, X., Ong, S. K., Nee, A. Y. C., Zhang, X. T., & Liu, W. J. (2011). Graonto: a graph-based approach for automatic construction of domain ontology. *Expert Systems with Applications*, *38*(9), 11958–11975. https://doi.org/10.1016/j.eswa.2011.03.090

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Huang, H., & Gartner, G. (2009). Using activity theory to identify relevant context parameters. In G. Gartner & K. Rehrl (Eds.), *Location Based Services and TeleCartography II* (pp. 35–45). Berlin, Heidelberg, Germany: Springer.

Hurley, C. (1998). The making and keeping of record: (1) what are finding aids for? *Archives and Manuscripts*, *21*(1), 58–77.

Hurley, C. (2005). Parallel provenance: (1) what, if anything, is archival description? *Archives and Manuscripts*, *33*(1), 110–145.

Hurley, C. (2009). *Hurley's common practice rules for the documentation of archives & other records*. Retrieved from https://www.descriptionguy.com/images/WEBSITE/hcpr-hurleys-common-practice-rules-2009.pdf

Hurley, C. (2012). Some things archivists do ... description & arrangement. Retrieved from http://www.descriptionguy.com/images/WEBSITE/some-things-archivists-do-arrangement&description.pdf

Hurley, C. (2017). Records in Context (RiC) 1.0 - Comments on First Draft (2016). Retrieved February 13, 2019, from http://www.descriptionguy.com/description.html#records_context

Hyvönen, E. (2012). *Publishing and using cultural heritage linked data on the semantic web* (Vol. 2). San Rafael, CA: Morgan & Claypool.

International Council of Archives. (2000). *ISAD(G): General international standard archival description; adopted by the Committee on Descriptive Standards, Stockholm, Sweden, 19-22 September 1999* (2. ed). Ottawa: International Council of Archives.

International Council on Archives. (2004). *International standard archival authority record for corporate bodies, persons, and families*. Ottawa: International Council on Archives. Retrieved from http://www.icacds.org.uk/eng/ISAAR%28CPF%292ed.pdf

International Council on Archives. (2007). *International standard for describing functions*. Ottawa: International Council on Archives. Retrieved from http://www.ica.org/download.php?id=1667

InterPARES Trust. (2016, December 11). InterPARES Trust responds to EGAD-RiC. Retrieved February 13, 2019, from https://interparestrust.com/2016/12/11/interpares-trust-responds-to-egad-ric/

Jenkinson, H. (1922). *A manual of archive administration including the problems of war archives and archive making*. Oxford, United Kingdom: The Clarendon Press.

Jick, T. D. (1979). Mixed qualitative and quantitative methods: triangulation in action. *Administrative Science Quarterly*, *24*(4), 602–611.

Jones, D., Bench-Capon, T., & Visser, P. (1998). Methodologies for ontology development. Retrieved from http://cgi.csc.liv.ac.uk/~tbc/publications/itknows.pdf

Jones, M. (2017, February 1). Comments on Records in Contexts. Retrieved February 13, 2019, from http://www.mikejonesonline.com/contextjunky/2017/02/02/comments-on-records-in-contexts/

Joy, C., Keane, A., & Corrigan, P. (2017). A landscape archive: methods for interaction design, preservation, access, and mapping—a case study. *Journal of Web Librarianship*, *11*(3–4), 209–219. https://doi.org/10.1080/19322909.2017.1378149

Kaenampornpan, M., & O'Neill, E. (2004). Modelling context: an activity theory approach. In P. Markopoulos, B. Eggen, E. Aarts, & J. L. Crowley (Eds.), *European Symposium on Ambient intelligence* (pp. 367–374). Berlin, Heidelberg, Germany: Springer.

Kalfatovic, M. R., Kapsalis, E., Spiess, K. P., Van Camp, A., & Edson, M. (2008). Smithsonian team Flickr: a library, archives, and museums collaboration in web 2.0 space. *Archival Science*, *8*(4), 267–277. https://doi.org/10.1007/s10502-009-9089-y

Katz, G., Pustejovsky, J., & Schilder, F. (2005). 05151 summary – annotating, extracting and reasoning about time and events. In G. Katz, J. Pustejovsky, & F. Schilder (Eds.), *Annotating, Extracting and Reasoning about Time and Events* (pp. 1–9). Dagstuhl, Germany: Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany.

Keneley, M., Potter, B., West, B., Cobbin, P., & Chang, S. (2016). Digitizing archival records: benefits and challenges for a large professional accounting association. *Archivaria*, *81*, 75–100.

Keshk, M., & Chambless, S. (2008). Model driven ontology: a new methodology for ontology development. In K. B. Lakey & D. Wijesekera (Eds.), *Proceedings of the Third International Ontology for the Intelligence Community Conference*. Fairfax, VA. Retrieved from http://ceur-ws.org/Vol-440/paper14.pdf

Khan, Z., & Keet, C. M. (2012). Onset: automated foundational ontology selection and explanation. In A. ten Teije, J. Völker, S. Handschuh, H. Stuckenschmidt, M. d'Acquin, A. Nikolov, … N. Hernandez (Eds.), *Knowledge engineering and knowledge management* (pp. 237–251).

Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-33876-2_22

Kim, J. (1993). *Supervenience and mind: selected philosophical essays*. New York, NY: Cambridge University Press.

Kofod-Petersen, A., & Cassens, J. (2006). Using activity theory to model context awareness. In T. R. Roth-Berghofer, S. Schulz, & D. B. Leake (Eds.), *Modeling and retrieval of context* (pp. 1–17). Berlin, Heidelberg, Germany: Springer Berlin Heidelberg.

Koivunen, M.-R., & Miller, E. (2011). W3c semantic web activity. Retrieved from http://www.w3.org/2001/12/semweb-fin/w3csw

Korstjens, I., & Moser, A. (2018). Series: Practical guidance to qualitative research. Part 4: Trustworthiness and publishing. *European Journal of General Practice*, *24*(1), 120–124. https://doi.org/10.1080/13814788.2017.1375092

Kowalski, R., & Sergot, M. (1986). A logic-based calculus of events. *New Generation Computing*, *4*, 67–75.

Krippendorff, K. (2004). *Content analysis: an introduction to its methodology* (2nd ed.). Sage Publications.

Krueger, R. A., & Casey, M. A. (2009). *Focus groups: a practical guide for applied research* (4th ed). Thousand Oaks, CA: Sage Publications.

Kumar, S. (2011). *Archives principles & practices*. Delhi, India: Gyan Publishing House. Retrieved from https://books.google.ca/books?id=fJOPXiLjF18C

Lack, R. (2006). The importance of user-centered design: exploring findings and methods. *Journal of Archival Organization*, *4*(1–2), 69–86. https://doi.org/10.1300/J201v04n01_05

Lagoze, C., & Hunter, J. (2001). The ABC ontology and model. In K. Oyama & H. Gotoda (Eds.), *Proceedings of the International Conference on Dublin Core and Metadata Applications 2001* (pp. 160–176). National Institute of Informatics, Tokyo, Japan.

Lagoze, C., & Hunter, J. (2002). The ABC ontology and model. *Journal of Digital Information, 2*(2). Retrieved from http://journals.tdl.org/jodi/index.php/jodi/article/view/44

Lebo, T., Sahoo, S., & McGuinness, D. (2013). Prov-o: the prov ontology. Retrieved August 8, 2018, from https://www.w3.org/TR/prov-o/#prov-o-at-a-glance

Lee, C. A. (2011). A framework for contextual information in digital collections. *Journal of Documentation*, *67*(1), 95–143. https://doi.org/10.1108/00220411111105470

Lemieux, V. L. (2014). Toward a “third order” archival interface: research notes on some theoretical and practical implications of visual explorations in the Canadian context of financial electronic records. *Archivaria*, *78*, 53–93.

Lemieux, V. L. (2015). Visual analytics, cognition and archival arrangement and description: studying archivists' cognitive tasks to leverage visual thinking for a sustainable archival future. *Archival Science*, *15*(1), 25–49. https://doi.org/10.1007/s10502-013-9212-y

Lewis, D. K. (1986). *On the plurality of worlds*. Oxford, United Kingdom: Blackwell Publishers.

Library of Congress. (2018). EAD: Encoded Archival Description (EAD official site, Library of Congress). Retrieved June 20, 2018, from http://www.loc.gov/ead/

Light, M., & Hyry, T. (2002). Colophons and annotations: new directions for the finding aid. *The American Archivist*, *65*(2), 216–230.

Lin, Y. (2008). *Semantic annotation for process models: facilitating process knowledge management via semantic interoperability*. Norwegian University of Science and Technology, Trondheim, Norway.

LOCAH Project. (2011). Locah project: creating archival linked data since 2010. Retrieved from http://archiveshub.ac.uk/locah/

López, M. F., Gómez-Pérez, A., Sierra, J. P., & Sierra, A. P. (1999). Building a chemical ontology using methontology and the ontology design environment. *IEEE Intelligent Systems*, *14*(1), 37–46.

Manola, F., & Miller, E. (2004). *Rdf primer*. Retrieved from http://www.w3.org/TR/2004/REC-rdf-primer-20040210/

Masolo, C., Borgo, S., Gangemi, A., Guarino, N., & Oltramari, A. (2003). *Wonderweb deliverable d18 : ontology library (final)*. Trento, Italy: Laboratory For Applied Ontology - ISTC-CNR. Retrieved from http://www.loa-cnr.it/Papers/D18.pdf

Matthews, B. (2005). *Semantic web technologies*. CCLRC Rutherford Appleton Laboratory. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.108.5427&rep=rep1&type=pdf

Mayring, P. (2000). Qualitative content analysis. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *1*(2). Retrieved from http://www.qualitative-research.net/index.php/fqs/article/view/1089

Mazzini, S., & Ricci, F. (2011). EAC-CPF ontology and linked archival data. In *Proceedings of the 1st international workshop on semantic digital archives* (Vol. 801, pp. 72–81). Berlin, Germany: CEUR workshop proceedings. Retrieved from http://ceur-ws.org/Vol-801/paper6.pdf

McCarthy, J., & Hayes, P. J. (1969). Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer, D. Michie, & M. Swann (Eds.), *Machine Intelligence 4* (pp. 463–502). Edinburgh, Scotland: Edinburgh University Press.

McCausland, S. (2011). A future without mediation? Online access, archivists, and the future of archival research. *Australian Academic & Research Libraries*, *42*(4), 309–319. https://doi.org/10.1080/00048623.2011.10722243

McGill University Archives. (2018). Madeleine Parent Fonds - Archival Collections Catalogue. Retrieved January 28, 2019, from https://archivalcollections.library.mcgill.ca/index.php/madeleine-parent-fonds

McKemmish, S., Acland, G., Ward, N., & Reed, B. (2006). Describing records in context in the continuum: the Australian recordkeeping metadata schema. *Archivaria*, *48*, 3–37.

Millar, L. (2002). The death of the fonds and the resurrection of provenance: archival context in space and time. *Archivaria*, *53*, 1–15.

Millar, L. (2010). *Archives: principles and practices*. New York, NY: Neal-Schuman Publishers.

Miller, F. (1990). *Arranging and describing archives and manuscripts*. Chicago, IL: Society of American Archivists.

Moreau, L., & Missier, P. (2013). PROV-DM: The PROV Data Model. Retrieved June 6, 2018, from https://www.w3.org/TR/2013/REC-prov-dm-20130430/#dfn-provenance

Morgan, D. (1997). *Focus groups as qualitative research* (2nd ed.). Thousand Oaks, CA: SAGE Publications. https://doi.org/10.4135/9781412984287

Muller, S., Feith, J. A., & Fruin, R. (2003). *Manual for the arrangement and description of archives: drawn up by direction of the netherlands association of archivists* (Translation of the 2nd ed.). Chicago, IL: Society of American Archivists.

Mwanza, D. (2002). *Towards an activity-oriented design model for HCI research and practice*. The Open University, Milton Keynes, United Kingdom.

Nardi, B. A. (1996). Studying context: a comparison of activity theory, situated action models, and distributed cognition. In *Context and consciousness: Activity theory and human–computer interaction.* (pp. 69–102). Cambridge, MA, US: The MIT Press.

National Archives of Australia. (2016). Australian Governments' Interactive Functions Thesaurus (AGIFT). Retrieved October 10, 2018, from https://data.naa.gov.au/def/agift.html

National Archives of Finland. (2003). *Finnish conceptual model for archival description (draft)*. National Archives of Finland. Retrieved from https://www.yumpu.com/en/document/view/36837292/the-finnish-conceptual-model-of-archival-description

Nesmith, T. (1999). Still fuzzy, but more accurate: some thoughts on the "ghosts" of archival theory. *Archivaria*, *47*, 136–150.

Nesmith, T. (2002). Seeing archives: postmodernism and the changing intellectual place of archives. *The American Archivist*, *65*(1), 24–41.

Nesmith, T. (2006). The concept of societal provenance and records of nineteenth-century Aboriginal–European relations in Western Canada: implications for archival theory and practice. *Archival Science*, *6*(3–4), 351–360. https://doi.org/10.1007/s10502-007-9043-9

Neuhaus, F., Grenon, P., & Smith, B. (2004). A formal theory of substances, qualities, and universals. In A. Varzi & L. Vieu (Eds.), *Formal Ontology in Information Systems: Proceedings of the Third International Conference (FOIS-2004)* (pp. 49–59). Amsterdam, Netherlands: IOS Press.

Neuman, W. L. (2003). *Social research methods: qualitative and quantitative approaches.* Boston, MA: Allyn & Bacon.

Nielsen, J. (1999). *Designing web usability: the practice of simplicity*. Thousand Oaks, CA: New Riders Publishing.

Nimer, C., & Daines, J. G. (2008). What Do You Mean It Doesn't Make Sense? Redesigning Finding Aids from the User's Perspective. *Journal of Archival Organization*, *6*(4), 216–232. https://doi.org/10.1080/15332740802533214

Nimer, C., & Daines, J. G. (2008). What do you mean it doesn't make sense? Redesigning finding aids from the user's perspective. *Journal of Archival Organization*, *6*(4), 216–232. https://doi.org/10.1080/15332740802533214

Niu, J. (2013). Provenance: crossing boundaries. *Archives and Manuscripts*, *41*(2), 105–115. https://doi.org/10.1080/01576895.2013.811426

Niu, J. (2014). Event-based archival information organization. *Archival Science*, *15*(3), 1–14. https://doi.org/10.1007/s10502-014-9222-4

Niu, J. (2015). Original order in the digital world. *Archives and Manuscripts*, *43*(1), 61–72. https://doi.org/10.1080/01576895.2014.958863

Niu, J. (2016). Linked data for archives. *Archivaria*, *82*, 83–110.

Noy, N. F., & McGuinness, D. L. (2001). *Ontology development 101: a guide to creating your first ontology* (Technical Report No. KSL-01-05). Stanford, California, USA: Stanford University. Retrieved from http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness-abstract.html

OCLC. (2018). Contentdm. Retrieved March 26, 2018, from https://www.oclc.org/en/contentdm.html

Oxford English Dictionary. (2014). Event, n. *Oxford English Online Dictionary*. Retrieved from http://www.oed.com/view/Entry/65287

Pan-Canadian Documentary Heritage Network (PCDHN). (2012). *Pan-Canadian Documentary Heritage Network (PCDHN) Linked Open Data (LOD) visualization "proof- of-concept." out of the trenches: linked open data of the First World War*. Retrieved from http://www.canadiana.ca/sites/pub.canadiana.ca/files/PCDHN%20Proof-of-concept_Final-Report-ENG_0_0.pdf

Panitch, J. M. (2001). *Special collections in ARL libraries: results of the 1998 survey: sponsored by the ARL Research Collections Committee*. Washington, DC: Association of Research Libraries.

Papajorgji, P., Pinet, F., Roussey, C., Brun, T., & Vigier, F. (2009). The Use of UML as a Tool for the Formalisation of Standards and the Design of Ontologies in Agriculture.

Park, J., Sung, K., & Moon, S. (2008). Developing Graduation Screen Ontology Based on the Methontology Approach. In *Proceeding of 2008 Fourth International Conference on Networked Computing and Advanced Information Management* (Vol. 2, pp. 375–380). Gyeongju, South Korea: IEEE. https://doi.org/10.1109/NCM.2008.215

Park, O. N. (2015). Development of linked data for archives in Korea. *D-Lib Magazine*, *21*(3/4). https://doi.org/DOI: 10.1045/march2015-park

Patel-Schneider, P. F., Hayes, P., & Horrocks, I. (2004). *Owl web ontology language semantics and abstract syntax*. Retrieved from http://www.w3.org/TR/owl-semantics

Pattuelli, M. C., Weller, C., & Szablya, G. (2011). Linked jazz: an exploratory pilot. In *International Conference on Dublin Core and Metadata Applications* (pp. 158–164). Hague. Retrieved from http://dcpapers.dublincore.org/pubs/article/view/3637

Pearce-Moses, R. (2005). *A glossary of archival and records terminology*. Chicago, IL: Society of American Archivists.

Pease, A. (2011). *Ontology: a practical guide*. Articulate Software Press. Retrieved from https://books.google.ca/books?id=Mvs3twAACAAJ

Pease, Adam, & Niles, I. (2002). IEEE standard upper ontology: a progress report. *The Knowledge Engineering Review*, *17*(1), 65–70.

Pease, Adam. (2018). Suggested Upper Merged Ontology (SUMO). Retrieved from http://www.adampease.org/OP/

Phytila, C. (2002). An analysis of the SUMO and description in unified modeling language. Retrieved from https://web.archive.org/web/20071012182612/http://ontology.teknowledge.com:80/Phytila/Phytila-SUMO.html

Piché, J.-S. (1998). Doing what's possible with what we've got: using the World Wide Web to integrate archival functions. *American Archivist*, *61*(1), 106–122.

Pickard, A. J. (2013). *Research methods in information* (2nd ed.). London, UK: Facet Publishing.

Piggott, M. (2012). 1 - Introduction: societal provenance. In M. Piggott (Ed.), *Archives and societal provenance* (pp. 1–12). Oxford, England: Chandos Publishing. https://doi.org/10.1016/B978-1-84334-712-5.50001-6

Pinet, F., Roussey, C., Brun, T., & Vigier, F. (2009). The Use of UML as a Tool for the Formalisation of Standards and the Design of Ontologies in Agriculture. In P. Papajorgji, *Advances in Modeling Agricultural Systems* (Vol. 25, pp. 131–147). Boston, MA: Springer US. https://doi.org/10.1007/978-0-387-75181-8_7

Pitti, D. V. (1999). Encoded archival description: an introduction and overview. *New Review of Information Networking*, *5*(1), 61–69. https://doi.org/10.1080/13614579909516936

Pitti, D. V. (2004). Creator description: encoded archival context. *Cataloging & Classification Quarterly*, *38*(3–4), 201–226. https://doi.org/10.1300/J104v38n03_16

Preisler, B. (1997). *A handbook of English grammar on functional principles* (2nd ed.). Aarhus, Denmark: Aarhus University Press.

Prom, C. (2004). User interactions with electronic finding aids in a controlled setting. *The American Archivist*, *67*(2), 234–268.

Pustejovsky, J., Castaño, J., Ingria, R., Saurí, R., Gaizauskas, R., Setzer, A., … Radev, D. (2003). Timeml: robust specification of event and temporal expressions in text. In M. T. Maybury (Ed.), *New Directions in Question Answering* (pp. 28–34). Menlo Park, CA: The AAAI Press.

Raimond, Y., & Abdallah, S. (2007). *The event ontology*. Retrieved from http://motools.sourceforge.net/event/event.html

Ranwez, S., Duthil, B., Sy, M. F., Montmain, J., Augereau, P., & Ranwez, V. (2013). How ontology based information retrieval systems may benefit from lexical text analysis. In A. Oltramari, P. Vossen, L. Qin, & E. Hovy (Eds.), *New Trends of Research in Ontologies and Lexical Resources: Ideas, Projects, Systems* (pp. 209–231). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-31782-8_11

Ren, Y., Parvizi, A., Mellish, C., Pan, J. Z., van Deemter, K., & Stevens, R. (2014). Towards Competency Question-Driven Ontology Authoring. In V. Presutti, C. d'Amato, F. Gandon, M. d'Aquin, S. Staab, & A. Tordai (Eds.), *The Semantic Web: Trends and Challenges* (pp. 752–767). Cham: Springer International Publishing.

Ren, Y., Parvizi, A., Mellish, C., Pan, J. Z., van Deemter, K., & Stevens, R. (2014). Towards competency question-driven ontology authoring. In V. Presutti, C. d'Amato, F. Gandon, M. d'Aquin, S. Staab, & A. Tordai (Eds.), *The Semantic Web: Trends and Challenges* (pp. 752–767). Cham: Springer International Publishing.

Report from Toronto Archival Context Meeting. (2001). Report from Toronto Archival Context Meeting. Retrieved from http://www.library.yale.edu/eac/torontoreport-final.htm

Rhee, H. L. (2012). Genres and genre repertoires of user and use information sources in U.S. state archival and records management appraisal practice. *Archival Science*, *12*(4), 461–483. https://doi.org/10.1007/s10502-012-9176-3

Rhee, H. L. (2015). Reflections on archival user studies. *Reference & User Services Quarterly*, *54*(4), 29–42.

Riffe, D., Lacy, S., & Fico, F. G. (1998). *Analyzing media messages: using quantitative content analysis in research*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Roberts, J., & Wilson, A. (2005). DCMI: DCMI agents working group. Retrieved October 10, 2018, from http://www.dublincore.org/groups/agents/

Rose, S., Spinks, N., & Canhoto, A. I. (2015). *Management research: applying the principles* (1 edition). London, UK: Routledge.

Rowat, T. (1993). The record and repository as a cultural form of expression. *Archivaria*, *36*, 198–204.

Rumbaugh, J., Jacobson, I., & Booch, G. (Eds.). (1999). *The unified modeling language reference manual*. Essex, UK: Addison-Wesley Longman Ltd.

Rumbaugh, J., Jacobson, I., & Booch, G. (Eds.). (1999). *The unified modeling language reference manual*. Essex, UK: Addison-Wesley Longman Ltd.

Sabourin, D. (2017), Madeleine Parent. In *The Canadian Encyclopedia*. Retrieved from https://www.thecanadianencyclopedia.ca/en/article/madeleine-parent

Saldaña, J. (2015). *The Coding Manual for Qualitative Researchers* (3rd edition). Los Angeles ; London: SAGE Publications Ltd.

Saurí, R., Goldberg, L., Verhagen, M., & Pustejovsky, J. (2009). *Annotating Events in English. TimeML Annotation Guidelines*. Brandeis University. Version TempEval-2010. Retrieved from http://www.timeml.org/tempeval2/tempeval2-trial/guidelines/EventGuidelines-050409.pdf

Saurí, R., Goldberg, L., Verhagen, M., & Pustejovsky, J. (2009). *Annotating events in English. Timeml annotation guidelines*. Brandeis University. Version TempEval-2010. Retrieved from http://www.timeml.org/tempeval2/tempeval2-trial/guidelines/EventGuidelines-050409.pdf

Sawsaa, A., & Lu, J. (2012). Building information science ontology (ois) with methontology and protégé. *Journal of Internet Technology and Secured Transactions (JITST)*, *1*(3/4), 100–109.

Scheir, W. (2005). First entry: report on a qualitative exploratory study of novice user experience with online finding aids. *Journal of Archival Organization*, *3*(4), 49–85. https://doi.org/10.1300/J201v03n04_04

Schellenberg, T. R. (1956). *Modern archives : principles and techniques*. Chicago, IL: University of Chicago Press.

Scherp, A., Franz, T., Saathoff, C., & Staab, S. (2009). F-a model of events based on the foundational ontology dolce+dns ultralight. In *Proceedings of the fifth international conference on Knowledge capture* (pp. 137–144). New York, NY: ACM.

Schreiber, G. (2005). Owl restrictions. Retrieved October 10, 2018, from https://www.cs.vu.nl/~guus/public/owl-restrictions/

Schreiber, G., & Raimond, Y. (2014). Rdf 1.1 primer. Retrieved June 20, 2018, from https://www.w3.org/TR/rdf11-primer/

Schutt, R. K. (2012). *Investigating the social world the process and practice of research* (7th ed.). Thousand Oaks, CA: SAGE Publications, Inc.

Schwartz, J. M., & Cook, T. (2002). Archives, records, and power: the making of modern memory. *Archival Science*, *2*(1–2), 1–19. https://doi.org/10.1007/BF02435628

Scott, P. J. (1966). The record group concept: a case for abandonment. *The American Archivist*, *29*(4), 493–504.

Semy, S. K., Pulvermacher, M. K., & Obrst, L. J. (2004). *Toward the use of an upper ontology for u.s. government and u.s. military domains: an evaluation* (p. 43). The MITRE Corporation.

Shaw, R. (2010). *Events and periods as concepts for organizing historical knowledge*. University of California, Berkeley.

Shepherd, E., & Yeo, G. (2003). *Managing records: a handbook of principles and practice*. London, England: Facet Publishing.

Singhal, A. (2012, May 16). Introducing the Knowledge Graph: things, not strings. Retrieved January 15, 2019, from https://www.blog.google/products/search/introducing-knowledge-graph-things-not/

Smith, J., Flowers, P., & Larkin, M. (2009). *Interpretative phenomenological analysis: theory, method and research* (1 edition). Los Angeles, CA: Sage Publications.

Society of American Archivists. (1999). Standards for archival description: a handbook. Retrieved from https://www.archivists.org/catalog/stds99/intro.html

Society of American Archivists. (2004). *Describing archives: a content standard* (1st ed.). Chicago, IL: Society of American Archivists.

Society of American Archivists. (2010). EAD (Encoded Archival Description) help pages--implementor listing. Retrieved October 10, 2018, from https://web.archive.org/web/20120412191412/http://www.archivists.org/saagroups/ead/implementors.html

Society of American Archivists. (2013). *Describing archives: a content standard* (2nd ed.). Chicago, IL: Society of American Archivists.

Society of American Archvists Working Group. (1989). Report of the working group on standards for archival description. *The American Archivist*, *52*(4), 440–461.

Spiro, L. (2009). *Archival management software: a report for the Council on Library and Information Resources*. Washington, D.C: Council on Library and Information Resources.

Staatsbibliothek zu Berlin. (2017). EAC-CPF homepage – encoded archival context for corporate bodies, persons, and families. Retrieved June 20, 2018, from http://eac.staatsbibliothek-berlin.de/

Stake, R. E. (1995). *The art of case study research*. Thousand Oaks, CA: SAGE Publications. Retrieved from https://books.google.ca/books?id=ApGdBx76b9kC

Stanford Center for Biomedical Informatics Research. (2016). Protégé. Retrieved April 27, 2018, from https://protege.stanford.edu/

Steinbruck, J. (1999). *The Yellowknife Journal*. Winnipeg, Canada: Nuage Editions.

Studer, R., Benjamins, V. R., & Fensel, D. (1998). Knowledge engineering: principles and methods. *Data Knowledge Engineering*, *25*(1–2), 161–197. https://doi.org/10.1016/S0169-023X(97)00056-6

Styler IV, W., Bethard, S., Finan, S., Palmer, M., Pradhan, S., de Groen, P., … Pustejovsky, J. (2014). Temporal annotation in the clinical domain. *Transactions of the Association for Computational Linguistics*, *2*, 143–154.

Suárez-Figueroa, M. C., Gómez-Pérez, A., & Villazón-Terrazas, B. (2009). How to write and use the ontology requirements specification document. In R. Meersman, T. Dillon, & P. Herrero (Eds.), *On the Move to Meaningful Internet Systems: OTM 2009* (pp. 966–982). Berlin, Heidelberg, Gemany: Springer Berlin Heidelberg.

Sure, Y., Staab, S., & Studer, R. (2004). On-to-knowledge methodology (otkm). In S. Staab & R. Studer (Eds.), *Handbook on Ontologies* (pp. 117–132). Berlin, Heidelberg, Gemany: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-24750-0_6

Sweeney, S. (2008). The ambiguous origins of the archival principle of “provenance.” *Libraries & the Cultural Record*, *43*(2), 193–213. https://doi.org/10.1353/lac.0.0017

Szary, R. V. (2006). Encoded archival context (eac) and archival description: rationale and background. *Journal of Archival Organization*, *3*(2–3), 217–227. https://doi.org/10.1300/J201v03n02_16

Talantikite, H. N., Aissani, D., & Boudjlida, N. (2009). Semantic annotations for web services discovery and composition. *Computer Standards & Interfaces*, *31*(6), 1108–1117. https://doi.org/10.1016/j.csi.2008.09.041

Tewksbury, R. (2009). Qualitative versus quantitative methods: understanding why qualitative methods are superior for criminology and criminal justice. *Journal of Theoretical & Philosophical Criminology*, *1*(1), 38–58.

The George W. Bush Presidential Library and Museum. (2018). Finding aids - the George W. Bush Presidential Library and Museum. Retrieved June 29, 2018, from https://www.georgewbushlibrary.smu.edu/Research/Finding-Aids

The InterPARES project team. (2001). The Interpares glossary. Retrieved from http://www.interpares.org/book/interpares_book_q_gloss.pdf

The InterPARES project team. (2018). Activity. Retrieved August 8, 2018, from http://www.interpares.org/ip3/ip3_terminology_db.cfm?term=967

The R Foundation. (n.d.). R: What is R? Retrieved September 20, 2018, from https://www.r-project.org/about.html

Thorsen, H. K., & Pattuelli, M. C. (2016). Linked opendata and the cultural heritage landscape. In *Linked data for cultural heritage* (pp. 1–22). Chicago, IL: Facet Publishing.

Thurman, A. C. (2005). Metadata standards for archival control: an introduction to ead and eac. *Cataloging & Classification Quarterly*, *40*(3–4), 183–212. https://doi.org/10.1300/J104v40n03_09

Tibbo, H. (2003). Primarily history in America: how us historians search for primary materials at the dawn of the digital age. *The American Archivist*, *66*(1), 9–50.

Tomlin, R. S. (1986). *Basic word order: functional principles*. London ; Wolfeboro, N.H: Croom Helm.

Trace, C. B., & Dillon, A. (2012). The evolution of the finding aid in the United States: from physical to digital document genre. *Archival Science*, *12*(4), 501–519. https://doi.org/10.1007/s10502-012-9190-5

Tsarkov, D., & Horrocks, I. (2006). FaCT++ Description Logic Reasoner: System Description. In U. Furbach & N. Shankar (Eds.), *Automated Reasoning: Third International Joint Conference, IJCAR 2006, Seattle, WA, USA, August 17-20, 2006. Proceedings* (pp. 292–297). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved from http://dx.doi.org/10.1007/11814771_26

Tsarkov, D., & Horrocks, I. (2006). Fact++ description logic reasoner: system description. In U. Furbach & N. Shankar (Eds.), *Automated Reasoning: Third International Joint Conference, IJCAR 2006, Seattle, WA, USA, August 17-20, 2006. Proceedings* (pp. 292–297). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved from http://dx.doi.org/10.1007/11814771_26

Ulin, P. R., Robinson, E. T., & Tolley, E. E. (2005). *Qualitative methods in public health: a field guide for applied research*. San Francisco, CA: John Wiley & Sons, Inc.

Uschold, Michael, & Gruninger, M. (2004). Ontologies and semantics for seamless connectivity. *SIGMOD Record*, *33*(4), 58–64.

Uschold, Mike, & Gruninger, M. (1996). Ontologies: principles, methods and applications. *The Knowledge Engineering Review*, *11*(2), 93–136. https://doi.org/10.1017/S0269888900007797

Uschold, Mike, & King, M. (1995). Towards a methodology for building ontologies. In *Workshop on Basic Ontological Issues in Knowledge Sharing, Held in Conjunction with IJCAI-95*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.55.5357

Uschold, Mike. (1996). Building ontologies: towards a unified methodology. In *In 16th Annual Conf. of the British Computer Society Specialist Group on Expert Systems*. Cambridge, England.

van Hage, W. R., Malaisé, V., Segers, R., Hollink, L., & Schreiber, G. (2011). Design and use of the simple event model (sem). *Journal of Web Semantics*, *9*(2), 128–136. https://doi.org/10.1016/j.websem.2011.03.003

W3C. (2007). Semantic web, and other technologies to watch. Retrieved from http://www.w3.org/2007/Talks/0130-sb-W3CTechSemWeb/#%2824%29

W3C. (2009). W3c semantic web FAQ. Retrieved January 30, 2011, from http://www.w3.org/2001/sw/SW-FAQ

W3C. (2015). Linked data. Retrieved September 21, 2018, from https://www.w3.org/standards/semanticweb/data

Walton, R. (2017). Looking for answers: a usability study of online finding aid navigation. *The American Archivist*, *80*(1), 30–52.

Weare, C., & Lin, W.-Y. (2000). Content analysis of the World Wide Web: opportunities and challenges. *Social Science Computer Review*, *18*(3), 272–292. https://doi.org/10.1177/089443930001800304

Weber, R. P. (1990). *Basic content analysis*. Thousand Oaks, CA: Sage Publications. Retrieved from https://books.google.ca/books?id=iv1oAAAAIAAJ

Weller, K. (2010). *Knowledge representation in the social semantic web*. Berlin, Germany: De Gruyter Saur.

White, M. D., & Marsh, E. E. (2006). Content analysis: a flexible methodology. *Library Trends*, *55*(1), 22–45. https://doi.org/10.1353/lib.2006.0053

Whittaker, B. M. (2007). DACS and RDA insights and questions from the new archival descriptive standard. *Libraries Resources and Technical Services*, *51*(2), 98–105.

Wildemuth, B. M. (2009). *Applications of Social Research Methods to Questions in Information and Library Science*. Westport, Conn: Libraries Unlimited.

Williams, C. (2006). *Managing archives: foundations, principles and practice*. Oxford, England: Chandos Publishing.

Winograd, T. (2001). Architectures for context. *Human–Computer Interaction*, *16*(2–4), 401–419. https://doi.org/10.1207/S15327051HCI16234_18

Woolcott, L., Payant, A., & Skindelien, S. (2016, October 25). Partnering for discoverability: knitting archival finding aids to digitized material using a low tech digital content linking process. The Code4Lib Journal. Retrieved from http://journal.code4lib.org/articles/11997

Yaco, S. (2008). It's complicated: barriers to EAD implementation. *American Archivist*, *71*(2), 456–475.

Yakel, E. (2003). Archival representation. *Archival Science*, *3*(1), 1–25. https://doi.org/10.1007/BF02438926

Yakel, E. (2004). Encoded archival description: are finding aids boundary spanners or barriers for users? *Journal of Archival Organization*, *2*(1–2), 63–77. https://doi.org/10.1300/J201v02n01_06

Yakel, E., & Kim, J. (2005). Adoption and diffusion of encoded archival description. *Journal of the American Society for Information Science and Technology*, *56*(13), 1427–1437. https://doi.org/10.1002/asi.20236

Yamagata-Lynch, L. C. (2010). *Activity systems analysis methods : understanding complex learning environments*. Boston, MA: Springer US. https://doi.org/10.1007/978-1-4419-6321-5

Yeo, G. (2007). Concepts of record (1): evidence, information, and persistent representations. *The American Archivist*, *70*(2), 315–343.

Yeo, G. (2008). Concepts of record (2): prototypes and boundary objects. *The American Archivist*, *71*(1), 118–143.

Yeo, G. (2012). Bringing things together: aggregate records in a digital age. *Archivaria*, *74*, 43–91.

Yeo, G. (2014). Contexts, original orders, and item-level orientation: responding creatively to users' needs and technological change. *Journal of Archival Organization*, *12*(3–4), 170–185. https://doi.org/10.1080/15332748.2015.1048626

Yin, R. K. (2003). *Case study research: design and methods*. Thousand Oaks, CA: SAGE Publications.

Yoose, B., & Perkins, J. (2013). The Linked Open Data Landscape in Libraries and Beyond. *Journal of Library Metadata*, *13*(2–3), 197-211. https://doi.org/10.1080/19386389.2013.826075

Zeng, M. L., & Qin, J. (2008). *Metadata* (1st ed.). New York, NY: Neal-Schuman Publishers.

Zhang, J., & Mauney, D. (2013). When archival description meets digital object metadata: a typological study of digital archival representation. *The American Archivist*, *76*(1), 174–195.

Zhang, Y., & Wildemuth, B. M. (2009). Qualitative analysis of content. In *Applications of Social Research Methods to Questions in Information and Library Science* (pp. 308–319). Westport, CT: Libraries Unlimited.

Zhou, X. (2007). Examining search functions of ead finding aids web sites. *Journal of Archival Organization*, *4*(3–4), 99–118. https://doi.org/10.1300/J201v04n03_06

APPENDIX I.

Archival descriptions coding guidelines

Introduction

This document describes tasks for identifying entities, relations, and events in archival descriptions. In general, there are three tasks (i.e., identification, annotation, and categorization). A set of guidelines explains the detailed processes how the tasks should be carried out.

Key terms used in this document are defined as follows:

Entity: a physical, digital, conceptual, or other kind of thing with some fixed aspects. It also "refers generally to all objects or agents of the system being analyzed" (Pearce-Moses, 2005, p. 148).

Event: "The (actual or contemplated) fact of anything happening; the occurrence of"; "Anything that happens, or is contemplated as happening; an incident, occurrence." (Oxford Online English Dictionary, 2014)

Relation: ways in which entities can be related to one another.

Role: A realizable entity the manifestation of which brings about some result or end that is not essential to a continuant in virtue of the kind of thing that it is but that can be served or participated in by that kind of continuant in some kinds of natural, social or institutional contexts.

Tasks

Coding archival descriptions is going through the description and marking entities, events, and relations. Each entity, event, and relation marked will be annotated for particular characteristics that provide more information and detail about each description. Then each new marked entity, event, and relation will be assigned a category if possible. If no category is available, a new category will be created and used. After each archival description, categories will be compared with each other as to whether categories need to be split, merged or change hierarchical levels. Therefore, there are three coding tasks: 1) identification, 2) annotation, and 3) categorization. After several tests, a five step coding procedure was developed as follows:

> **Step 1**. Identifying and locating events described by verbs and nominalized verbs
> **Step 2**. Identifying other phrases associated with each verb
> **Step 3**. Marking appropriate spans of text to represent the phrases
> **Step 4**. Assigning appropriate entities to the phrases
> **Step 5**. Assigning appropriate event types to the events and linking events with related entities

The three coding tasks are divided into smaller steps in the coding procedure. The core of Step 1 is to identify events. Step 2 is to identify associated entities. Step 3 and 4 are for marking. The core of Step 4 and 5 is to categorize entities and events. The five steps are explained with examples in following sections.

Step 1. Identifying and locating events described by verbs and nominalized verbs

An event is anything that is relevant to the creator's activities. More specifically, anything that would show up on a detailed timeline of a creator's history would be considered an event. So, formation of an organization would certainly appear on such a timeline of the organization. But temporally span-less entities like people, organizations, or non-anatomical objects (car) will never be events. In this document, events denoted by verbs or nouns are considered.

Events denoted by Verbs

The general approach is to check verbs and only the verb itself is considered as an event. In the sentences below, the verbal event expressions are indicated in bold face and between square brackets.

(1)	a. In 1889, Acadia University [**dissolved]** its three existing athletic clubs and **[formed]** the Acadia Amateur Athletic Association. b. It [**was assessioned**] on 26 Novermber 2001. c. He [**brought**] us various records. d. Rev. Alfred Chipman [**was born**] in 1834 at Pleasant Valley (now Berwick), Nova Scotia. e. She [**taught**] for one year at the Ladies Collegiate Institute of Worcheste.

There are a number of verb constructions including simple verbal phrases (VPs), complex VPs, phrasal verbs, and idioms. The general rule for events denoted by verbs is to identify the whole verbal expressions in each sentence.

Simple VPs. The event tag marks the verbal head.

(2)	a. She [**taught**] for one year at the Ladies Collegiate Institute of Worcheste. b. He [**graduated]** from Harvard University in 1996.

Complex VPs. The verbal head is accompanied by auxiliaries and related particles. Only the verbal head needs to be marked. In the following examples, the span of the verbal phrases is underlined, whereas the event is marked in bold face.

(3)	a. Israel [has been scrambling to **buy]** more masks abroad. b. Kaufman [did not **disclose**] details of the deal. c. The private sector [could **establish]** a private agency.

Phrasal verbs. Phrasal verbs are also known as verb-particle constructions.

(4)	a. Additional distribution centers [would be **set** up] next week. b. He [**sold** out] his shares to William Henry Cobb.

Idioms. Only the main verbal predicate will be marked as event.

(5)	a. Even more hard drives [**kick** the bucket].

Events denoted by Nouns

Event-denoting nouns often have a similar but different form to their related verbs (e.g., retirement – retire, analysis – analyze).

Event-denoting nouns acting as prenominal modifiers will **NEVER** be annotated as events but may be marked as features. See the following examples:

(6)	a. Many of the same reactions occur in a [**panic** attack]. b. The likely shape of the [**leadership** contest] emerged yesterday as Labour began an inquest on its [**election** defeat].

Other common cases involve the use of present participle forms. They will **not** be annotated as events.

(7)	a. the **waiting** room b. **knitting** needle c. **shooting** match d. **drinking** problem

All other event-denoting nouns will be marked up as events **ALWAYS**. For example:

a. James L. G. Fitz Patrick also donated several of his research publications to the College Archives prior to his [**retirement]** in 1976.
b. The economic [**chokehold]** appears to be working.
c. The men explained that that was an [**attack]** by masked individuals.
d. The financial [**assistance]** from the World Bank and the International Monetary Fund are not helping.

Only the head nouns are marked up.

Note: ALL verbs and nominalised verbs in the sentence should be annotated.

Once events have been identified, the next step is to identify other phrases associated with each verb.

Step 2. Identifying other phrases associated with each verb

This section addresses what to mark as the span, or extent, of the event tag. The pre-defined categories are "object, agent, occurrence, purpose, time, place, form of expression, concept/abstraction, and relationship" (Lee, 2011, p. 95). In general, this step is to identify phrases that may fall into the categories. If some phrases do not fit into the ten categories, a new category should be created for the phrases. The definition of the nine categories (Lee, 2011, 106) is listed as follows:

Object	A bounded discrete entity that can be characterized as having one or more properties or states; persist across multiple points in time and place; be uniquely identified; interact with other objects; and be acted upon by an agent
Agent	An entity that can carry out actions

Occurrence	A characterization, for a given span of times and places, of either the state of a set of entities or their interaction(s)
Purpose	Mandate, norms, values, intention, rules, standards, virtues, or functions to which agents can advance or with which they can conform; attempt to advance or conform; hope to advance or conform; or perceive/expect entities (or sets of entities) to advance or conform
Time	"A limited stretch or space of continued existence, as the interval between two successive events or acts, or the period through which an action, condition, or state continues" (Oxford English Dictionary, 1989)
Place	A designated point or region in space
Form of expression	A particular way of expressing ideas or information
Concept or Abstraction	Ideas or other individually/socially recognized "properties or qualities as distinguished from any particular embodiment of the properties/qualities in a physical medium" (Standard Upper Merged Ontology)
Relationship	An association between two or more entities (or classes of entities), which cannot be reduced to or adequately expressed as a property of the entities (or classes of entities) themselves

Table I.1. Nine classes of contextual entities (Lee, 2011, p. 106)

In order to make it easier to identify entities, the following template (see Table I.2) is designed for such purposes. The template is straightforward and can apply to a sentence.

Sentence #	Sentence text		
	Entity		**Text**
	Occurrence		
	Agent		
	Concept/Abstraction		
	Form of expression		
	Object		
	Place		
	Purpose		
	Time		
	Relation		
	New Category		

Table I.2. Coding template

For example, the following shows a sentence can be divided into smaller phrases using the template.

a. James L. G. Fitz Patrick also donated several of his research publications to the College Archives prior to his retirement in 1976.

Sentence #	James L. G. Fitz Patrick also donated several of his research publications to the College Archives prior to his retirement in 1976.		
	Entity		**Text**
	Occurrence		*donated* *retirement*
	Agent		*James L. G. Fitz Patrick* *the College Archives*
	Concept/Abstraction		
	Form of expression		*research publications*
	Object		
	Place		
	Purpose		
	Time		1976
	Relation		
	New Category		

Table I.3. A coding example

In terms of entities, the general approach is to check nouns and nouns phrases, and only the nouns and noun phrases themselves are considered entities. In the sentences below, the noun expressions are indicated in bold face and between square brackets.

(8)	a. In 1889, [**Acadia University**] dissolved its three existing [**athletic clubs**] and formed [**the Acadia Amateur Athletic Association**]. b. It was assessioned on 26 Novermber 2001. c. He brought us various [**records**]. d. Rev. [**Alfred Chipman**] was born in 1834 at Pleasant Valley (now Berwick), Nova Scotia. e. She taught for one year at [**the Ladies Collegiate Institute of Worcheste**].

In sentence 8a., Acadia University, athletic clubs, and the Acadia Amateur Athletic Association are three entities identified. They can be considered as instances of concept "Organization". In sentence 8d., Alfred Chipman is a person's name that can be an individual of the concept “Person”. In sentence 8e., the Ladies Collegiate Institute of Worcheste is an instance of the concept “Organization”.

In terms of time or date, the general rule is to include the completed time or date as possible. Do **not** include propositions.

(9)	a. In [**1889**], Acadia University dissolved its three existing athletic clubs and formed the Acadia Amateur Athletic Association. f. It was assessioned on [**26 Novermber 2001**]. g. He brought us various records. h. Rev. Alfred Chipman was born in [**1834**] at Pleasant Valley (now Berwick), Nova Scotia. i. She taught for [**one year**] at [the Ladies Collegiate Institute of Worcheste].

Step 3. Marking appropriate spans of text to represent the phrases
Before beginning an annotation, it is important to read through the complete sentence again, concentrating on the event denoted by the selected verb and trying to locate the phrases in the sentence that correspond to categories in the event. This step is to ensure that all associated entities of the event are located, and to ensure that the texts are properly marked.

The procedure is to start with the event and then find the associated entities. Take Table I. 3 as an example, there are two events (i.e., donated and retirement). The donation event occurred before 1976. However, in Table I.3, only “1976” is identified. The time entity of the donation event should be “Prior to 1976” and that of the retirement event is “1976”. For the sake of simplicity, the span of the time entity is “prior to his retirement in 1976”.

Sentence #	James L. G. Fitz Patrick also donated several of his research publications to the College Archives prior to his retirement in 1976.		
	Entity		**Text**
	Occurrence		*donated* *retirement*
	Agent		*James L. G. Fitz Patrick* *the College Archives*

	Concept/Abstraction		
	Form of expression		*research publications*
	Object		
	Place		
	Purpose		
	Time		**Prior to 1976 – donation event** **1976 – retirement event**
	Relation		
	New Category		

Table I.4. A coding example - modified

Step 4. Assigning appropriate entities to the phrases

This step is to assign each phrase of the event with a proper category. Table I.3 and Table I.4 have shown that the spans of text have been associated with the nine categories of entities. Step 1, 2, and 3 focused on identification of events and other entities and the proper spans of text that are associated with the entities. This step is to look at the text and evaluate whether the text is properly categorized. If yes, whether the nine categories are suitable and whether a specific sub-category is needed. If not, whether a new category is needed.

For example, in Table I.3 and Table I.4, there are no sub-categories. The events can be further categorized into a type of archival event (i.e., donation event) and life event (i.e., retirement event). In terms of *Agent,* there are two sub-categories (i.e., *Person* and Corporate Body).

Sentence #	James L. G. Fitz Patrick also donated several of his research publications to the College Archives prior to his retirement in 1976.		
	Entity		**Text**
	Occurrence	**Donation event** **Life event**	*donated* *retirement*
	Agent	**Person** **Corporate body**	*James L. G. Fitz Patrick* *the College Archives*
	Concept/Abstraction		
	Form of expression	**Information artifact**	*research publications*
	Object		
	Place		
	Purpose		
	Time		Prior to 1976 – donation event

			1976 – retirement event
	Relation		
	New Category		

Table I.5. A coding example –sub-categories

Note: All associated phrases of the event within the same sentence should to be annotated.

REMEMBER:

- Spans should normally consist of complete chunks (single chunks wherever possible)
- Short entity names are to be favoured over longer names or characterizations, if both are present within the sentence
- Descriptive information about entities should not be included within the span
- Where a variable consists of a list of entities, the span should consist of all items in the list, excluding commas and conjunctions etc.
- LOCATION and TEMPORAL spans should include the preposition that precedes them (e.g. in, after etc.)
- Assign an appropriate semantic role to the marked phrase, or UNDERSPECIFIED if none of the roles in the current set seems appropriate. In the case, it is important to include a comment that explains the perceived function of the phrase in the event.

Step 5. Assigning appropriate event types to the events and linking events with related entities

This step is to concentrate on further categorizing events. Events can be categorized by themes and by their relationship with states. In terms of states, events are grouped into the following groups: ExperienceEvent, TransferEvent, ProduceEvent, ExistenceChangeEvent, and ChangeInPossessionEvent. For example, the donation event is a type of ChangeInPosessionEvent (TransferEvent is about agent from one place to another). The retirement event is a type of ExperienceEvent.

Sentence #	James L. G. Fitz Patrick also donated several of his research publications to the College Archives prior to his retirement in 1976.		
	Entity		**Text**
	Occurrence	**Donation event** **Life event**	*donated* *retirement*
	Agent	**Person** **Corporate body**	*James L. G. Fitz Patrick* *the College Archives*
	Concept/Abstraction		
	Form of expression	**Information artifact**	*research publications*
	Object		
	Place		
	Purpose		
	Time		Prior to 1976 – donation event 1976 – retirement event
	Relation		
	New Category		

Table I.6. A coding example – event categories

The events are categorized in two steps. First, it is to fit them the five categories. Second, it is to classify them by theme (ArchivalEvent, MilitaryEvent, LifeEvent, etc.).

APPENDIX II.
Letter of invitation

Lakehead UNIVERSITY

Library
t: (807) 343-8251 f: (807) 343-8007
e: qzou@lakeheadu.ca

Representation of Archival Descriptions on the Semantic Web

I am inviting you to participate in a doctoral thesis study entitled "Representation of Archival Descriptions on the Semantic Web".

The purpose of the study is to explore how archival descriptions can be represented and re-organized using semantic web technologies with respect to the cluster of archival principles of provenance and context. You will participate in a focus group study that is the third stage of the study in order to evaluate the usefulness and applicability of the proposed event ontology in archival descriptions.

As a participant, you will be asked to evaluate the usefulness of representations of the current archival descriptions and event ontology-based model, answer questions and your opinions about archival descriptions and discuss with other participants as a group. It will take approximately 40 to 60 minutes of your time. At the end of the focus group session, you will receive a $100 for your time and contributions.

Your cooperation in serving as a participant will be both important and valuable. The information you provide will be valuable sources for this research.

There is no anticipated risk to you if you choose to take part. Your participation is entirely voluntary, and you have the authority to refuse to answer any questions and incomplete any tasks with no penalty. At any time, you may choose to withdraw from the research without penalty.

The Information you provide will be coded and used only for the purposes of this study. Any information that is obtained and that can be related to you will be kept confidential. The data will be kept on a password protected secure server at McGill University after the research is complete. All data from this study can only be accessed by the investigator and his dissertation committee.

Please contact me by email (qzou@lakeheadu.ca). If you are interested in participating in this study, or if you have any questions. Please feel free to forward this invitation to someone else you think might be interested in this research.

Best regards,

Qing Zou
Doctoral Candidate and Principal Investigator
Library
Lakehead University
955 Oliver Rd.
Thunder Bay, ON P7B 5E1
Tel.: (807) 343-8251
E-mail: qzou@lakeheadu.ca

Eun G. Park
Associate Professor and Faculty Advisor
School of Information Studies
McGill University
3661 Peel Street
Montreal, QC H3A 1X1
Tel.: (514) 398-3364
E-mail: eun.park@mcgill.ca

APPENDIX III.

Sample consent form

Representation of Archival Descriptions on the Semantic Web

You are invited to participate in a research study being conducted by Qing Zou, doctoral candidate at the School of Information Studies, McGill University, Montreal as a part of his dissertation research.

The purpose of the study is to explore how archival descriptions can be represented and re-organized using semantic web technologies with respect to the cluster of archival principles of provenance and context. You will participate in a focus group study that is the third stage of the study in order to evaluate the usefulness and applicability of the proposed event ontology in archival descriptions.

As a participant, you will be asked to evaluate the usefulness of representations of the current archival descriptions and event ontology-based model and discuss with other participants as a group. It will take approximately 40 to 60 minutes of your time. The focus group will be taken place in Room 310 at School of Information Studies, McGill University. During the session, the investigator will ask you a series questions relating to how archival descriptions are organized, what you think of the current archival description practices, and what are your opinions of the proposed event ontology-based model. The discussion will be audio recorded.

You will have the opportunity to express your thoughts on archival descriptions and you will learn more about ontologies, representation of archival descriptions, and Semantic Web. There is no anticipated risk to you if you choose to take part. Your participation is entirely voluntary, and you have the authority to refuse to answer any questions and incomplete any tasks with no penalty. At any time, you may choose to withdraw from the research without penalty.

All participants will receive a $100 cash for your time and contributions. If you choose to withdraw from the research after the session, you will not be asked to return the gift card.

Personal identifiable information will be removed as possible. Any Information you provide will be coded and used only for the purposes of this study. All data from this study will be encrypted and stored on a password protected secure server at McGill University for five years and can only be accessed by the researcher and advisor Professor Park. Any other persons must get your permission to access information you provide. The findings of the study are available to you at your request once the study is completed. The result of this study will be presented in a doctoral thesis and may appear in peer-reviewed scholarly journals. The information that may directly associate with you will not be revealed.

You will have the right to review your information before personal identifiable information being removed. You can withdraw your permission to use your information at any time. You also have the right to review the recordings made for this study and edit and/or erase any information with which you are not comfortable. Anything you say will only be attributed to you with your permission. The researcher will keep the recordings in a secure place and the recordings will be only accessed by the researcher and his advisor duration of the study. After the research is over, the recordings will be encrypted and kept in a password protected secure server at McGill following McGill's policy regarding audio recordings. Any information that is obtained and that can be related to you will remain confidential.

If you have any questions or concerns about this study, you are welcome to contact me at qing.zou2@mail.mcgill.ca or my advisor, Professor Eun Park, School of Information Studies, McGill University at (514) 398-3364 or eun.park@mcgill.ca.

If you have any questions or concerns regarding your rights and welfare as a participant in this research study, please contact the McGill Ethics Manager at (514) 398-6831 or Lynda.mcneil@mcgill.ca.

I have read the above information and I agree to participate in this study. I will receive a signed and dated copy of this form at the end of the focus group session.
I agree for my voice to be digitally recorded. ____ YES _____ NO

______________________________	______________________________
Signature of the participant	Signature of the principal investigator
______________________________	______________________________
Date	Date

APPENDIX IV.

Sample questionnaire

Please fill in the circle beside the response that best describes you or your use of archival descriptions.

1. Have you taken any courses in archival sciences (e.g. archival, records management, etc.) ?
    - o None
    - o 1 course
    - o 2 courses
    - o Other ______________________

2. Have you taken any metadata related courses (e.g. metadata, organization of information, etc.) ?
    - o None
    - o 1 course
    - o 2 courses
    - o Other _____________________

3. Have you visited archives?
    - o No
    - o Yes,

    if Yes, how often?
        - o 1 time/per month,
        - o 2 times/per months,
        - o Other _________

4. Have you used archival descriptions before?
    - o Yes

    if Yes, how often?
        - o 1 time/per month,
        - o 2 times/per months,
        - o Other _________
    - o No

5. Have you created archival descriptions before?
    - o No
    - o Yes,

    if Yes, how often?
        - o 1 time/per month,
        - o 2 times/per months,
        - o Other _________

APPENDIX V.

Three sample archival descriptions for focus group discussions

Archival Description I

Current Practice:

In 1889, Acadia University dissolved its three existing athletic clubs (football, baseball, and cricket) and formed the Acadia Amateur Athletic Association (A.A.A.A.).
Its initial mandate stated that it was "to promote an interest in the physical development of the students by means of healthy, vigorous and entertaining games, and to keep in condition a campus well-appointed for this purpose" (Acadia Athenaeum, Nov. 1894).
J. R. Herbin, the main force behind its formation, was appointed its first president.
…
It has not been determined with certainty when the A.A.A.A. was dissolved, but it was most likely about the 1969/70 school year as it last appears in the Acadia University yearbook in 1968.

Proposed Model

In 1889, Acadia University dissolved its three existing athletic clubs (football, baseball, and cricket) and formed the Acadia Amateur Athletic Association (A.A.A.A.). Its initial mandate stated that it was "to promote an interest in the physical development of the students by means of healthy, vigorous and entertaining games, and to keep in condition a campus well-appointed for this purpose" (Acadia Athenaeum, Nov. 1894). J. R. Herbin, the main force behind its formation, was appointed its first president. … It has not been determined with certainty when the A.A.A.A. was dissolved, but it was most likely about the 1969/70 school year as it last appears in the Acadia University yearbook in 1968.	EndOrgEvent: • 1889 • Acadia University athletic clubs (football, baseball, and cricket) StartOrgEvent: • 1889 • The Acadia Amateur Athletic Association BioEvent: • Inauguration • 1889 • first president • A.A.A.A. EndOrgEvent: • A.A.A.A • 1969/70 (?)

Archival Description II

Current Practice:
Alice Theodosia Shaw was born 21 June 1832, the daughter of Isaiah Shaw and Sarah Lyons.
At the age of 22 years she traveled to Massachusetts to attend the Mount Holyoke Seminary and after three years of study graduated in August 1857 with a MA degree.
She then taught for one year at the Ladies Collegiate Institute of Worchester, MA, returning to Nova Scotia in 1858.
…
Alfred and Alice were married at River Philip, Nova Scotia on 28 October 1862.
They had three children: Fred McCullock who was born in 1865 at Pictou, Nova Scotia and who died in Florida in 1892; Alvah Hovey who was born in 1867 in Stewiake, Nova Scotia (donor of these records); and Chaloner Oaken, born 1871 at Sydney, Nova Scotia.
Alice and Alfred retired to Berwick, NS in 1899.
Alfred died 24 April 1918, at Berwick, while Alice died at the age of 88, on 18 January 1921, also at Berwick.

Proposed Model:

<table>
<tr>
<td>Alice Theodosia Shaw was born 21 June 1832, the daughter of Isaiah Shaw and Sarah Lyons.
At the age of 22 years she traveled to Massachusetts to attend the Mount Holyoke Seminary and after three years of study graduated in August 1857 with a MA degree.
She then taught for one year at the Ladies Collegiate Institute of Worchester, MA, returning to Nova Scotia in 1858.
…
Alfred and Alice were married at River Philip, Nova Scotia on 28 October 1862.
They had three children: Fred McCullock who was born in 1865 at Pictou, Nova Scotia and who died in Florida in 1892; Alvah Hovey who was born in 1867 in Stewiake, Nova Scotia (donor of these records); and Chaloner Oaken, born 1871 at Sydney, Nova Scotia.
Alice and Alfred retired to Berwick, NS in 1899.
Alfred died 24 April 1918, at Berwick, while Alice died at the age of 88, on 18 January 1921, also at Berwick.</td>
<td>BioEvent:
• Birth (#1)
• 21 June 1832
• Isaiah Shaw
• Sarah Lyons
TravelEvent: (#2)
• At the age of 22
• Massachusetts
BioEvent:
• Graduation (#3)
• Mount Holyoke Seminary
• August 1857
• MA degree
BioEvent:
• Teach (#3)
• Ladies Collegiate Institute of Worchester, MA
• 1857-58
TravelEvent: (#4)
• 1858
• Nova Scotia
BioEvent:
• Marriage (#5)
• Alfred
• Alice
• River Philip, Nova Scotia
• 28 October 1862.
BioEvent:
• Birth (#6)
• Fred McCullock
• 1865
• Pictou, Nova Scotia</td>
</tr>
</table>

	• Death (#7) • Florida • 1892 • Birth (#8) • Alvah Hovey • 1867 • Stewiake, Nova Scotia • Donor • Birth (#9) • Chaloner Oaken • 1871 • Sydney, Nova Scotia. BioEvent: • Retirement (#10) • Alfred and Alice • 1899 • Berwick, NS • Death (#11) • Alfred • 24 April 1918 • Berwick • Death (#12) • Alice • 18 January 1921 • Berwick

Temporal Relations (chronological order):
#1-> #2-> #3->#4->#5->#6->#8->#9->#10->#12

#6->#7->#11

Archival Description III

Current practice:

This collection consists of 23 letters to and from 2nd Lieutenant (Confederate States Army) W.E. Johnson, Jr. dating from March 29, 1864 through December 28, 1864.

The majority of the letters were written by Johnson to his wife Ann Johnson or his father W.E. Johnson, Sr.

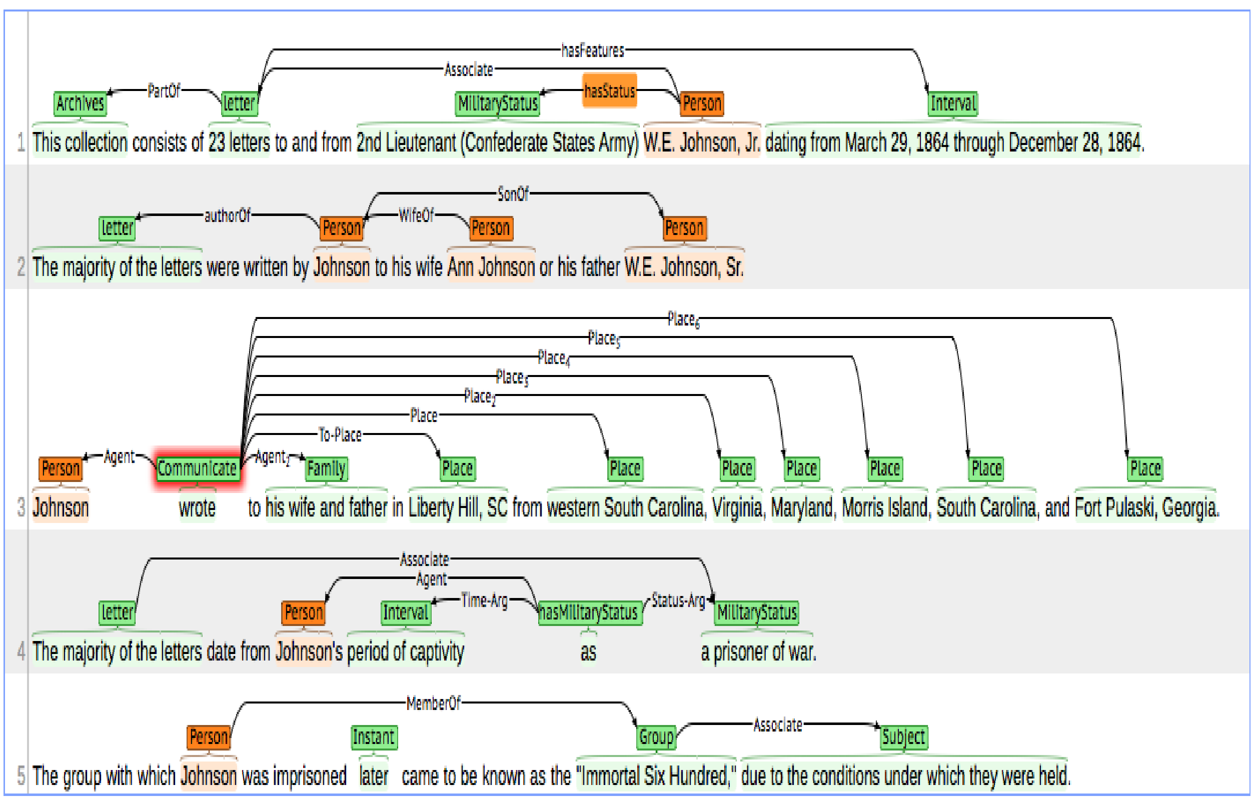
Johnson wrote to his wife and father in Liberty Hill, SC from western South Carolina, Virginia, Maryland, Morris Island, South Carolina, and Fort Pulaski, Georgia.

The majority of the letters date from Johnson's period of captivity as a prisoner of war.

The group with which Johnson was imprisoned later came to be known as the "Immortal Six Hundred," due to the conditions under which they were held.

Proposed Model:

This collection consists of 23 letters to and from 2nd Lieutenant (Confederate States Army) W.E. Johnson, Jr. dating from March 29, 1864 through December 28, 1864. The majority of the letters were written by Johnson to his wife Ann Johnson or his father W.E. Johnson, Sr. Johnson wrote to his wife and father in Liberty Hill, SC from western South Carolina, Virginia, Maryland, Morris Island, South Carolina, and Fort Pulaski, Georgia. The majority of the letters date from Johnson's period of captivity as a prisoner of war. The group with which Johnson was imprisoned later came to be known as the "Immortal Six Hundred," due to the conditions under which they were held.	See the following figure



APPENDIX VI.

Focus group questions

1. Do you think the current archival descriptions are helpful to fulfill your needs as a user?

2. If you are a professional or researcher, do you think the current archival descriptions fulfill the needs? Please specific advantages and disadvantages of using the current archival descriptions.

3. Based on the three archival descriptions you have seen, do you think the event ontology model is helpful to fulfill your needs? Please specific advantages and disadvantages of using the event ontology model.

4. Will you use the event model in future at your work place? Why?

5. If you have any other issues or comments on the current archival descriptions and event model, please address them.

APPENDIX VII.

McGill University Research Ethics Board approval form

McGill University
ETHICS REVIEW
RENEWAL REQUEST/STUDY CLOSURE FORM

Continuing review of research involving humans requires, at a minimum, the submission of an annual status report to the REB. This form must be completed to request renewal of ethics approval. If a renewal is not received before the expiry date, the project is not considered to be approved and no further research activity may be conducted. When a project has been completed, this form can also be used to officially close the study. To avoid expired approvals and, in the case of funded projects, the freezing of funds, this form should be returned 2-3 weeks before the current approval expires.

REB File #: 70-0814
Project Title: Representation of Archival Description on the Semantic Web
Principal Investigator: Qing Zou
Email: qing.zou2@mail.mcgill.ca
Faculty Supervisor (if PI is a student): Prof. E. Park

1. Were there any significant changes made to this research project that have any ethical implications that have not already been reported to the REB? ____YES __√_NO
   If yes, complete an amendment form indicating these changes and attach to this form.

2. Are there any ethical concerns that arose during the course of this research? ___ YES __√_ NO
   If yes, please describe.

3. Have any participants experienced any unanticipated issues or adverse events in connection with this research project? ___ YES __√_ NO
   If yes, please describe.

4. Is this a funded study? ___ YES _√__ NO
   If yes, list the agency name and project title and the Principal Investigator of the award if not yourself. This information is necessary to ensure compliance with agency requirements and that there is no interruption in funds.

5. Does this project require REB approval from another Institution/Board? _√__YES ___ NO
   If yes, and the project is continuing, attach a copy of the current approval.

Principal Investigator Signature: Qing Zou Date: __July 25, 2017______

Faculty Supervisor Signature: Eun Gyo Park Date:__ July 27, 2017______
(if PI is a student)

____ Check here if the **study is to be closed** and continuing ethics approval is no longer required. A study can be closed when all data collection has been completed and there will be no further contact with participants.

__√__ Check here if this is a **request for renewal** of ethics approval.

For Administrative Use **REB:** _X_REB-I ___REB-II ____REB-III

_X_ Delegated Review ____ Full Review

Signature of REB Chair or designate: L McNeil Date: Aug. 3, 2017

Approval Renewal Period: Aug. 3, 2017 to Aug. 2, 2018

Submit by email to lynda.mcneil@mcgill.ca. REB Office: James Administration Building, 845 Sherbrooke Street West suite 429, Mtl., QC H3A0G4; tel: 514-398-6831/6193; fax: 514-398-4644;www.mcgill.ca/research/researchers/compliance/human
(December 2014)

APPENDIX VIII.

Lakehead University Research Ethics Board approval letter

Lakehead UNIVERSITY

Research Ethics Board
t: (807) 343-8283
research@lakeheadu.ca

October 22, 2014

Principal Investigator: Eun G. Park
Co-Investigator: Qing Zou
External
McGill University

Dear Eun G. Park:

Re: REB Project #: 071 14-15 / Romeo File No: 1464108
Granting Agency: N/A
Granting Agency Project #:N/A

On behalf of the Research Ethics Board, I am pleased to grant ethical approval to your research project titled, "Representation of Archival Descriptions on Semantic Web".

Ethics approval is valid until October 23rd, 2015. Please submit a Request for Renewal form to the Office of Research Services by September 23rd, 2015 if your research involving human subjects will continue for longer than one year. A Final Report must be submitted promptly upon completion of the project. Research Ethics Board forms are available at: https://www.lakeheadu.ca/research-and-innovation/forms

During the course of the study, any modifications to the protocol or forms must not be initiated without prior written approval from the REB. You must promptly notify the REB of any adverse events that may occur.

Completed reports and correspondence may be directed to:

Research Ethics Board
c/o Office of Research Services
Lakehead University
955 Oliver Road
Thunder Bay, ON P7B 5E1
Fax: (807) 346-7749

Best wishes for a successful research project.

Sincerely,

Dr. Lori Chambers
Chair, Research Ethics Board

955 Oliver Road, Thunder Bay, ON, Canada, P7B 5E1 lakeheadu.ca

APPENDIX IX.

Information artifacts

Document
……Legal Document
…………Affidavits
…………Appeal
…………Case file
…………Judgement
…………Juristiction
……Certificate
…………Degree
………………BA
………………MA
………………PhD
……Architecture drawing
……Cartographic plan
……Correspondence
……Description
…………Finding aids
………………Archival Description
……Diary
……Ephemera
……Essay
……Letter
……Manuscript
……Map
……Memoradum
……Minutes
……Newspaper clipping
……Note
……Photograph
……Postcard
……Report
Publication
……Non-scholarly Publication
……Scholarly Publication

APPENDIX X.

A list of events

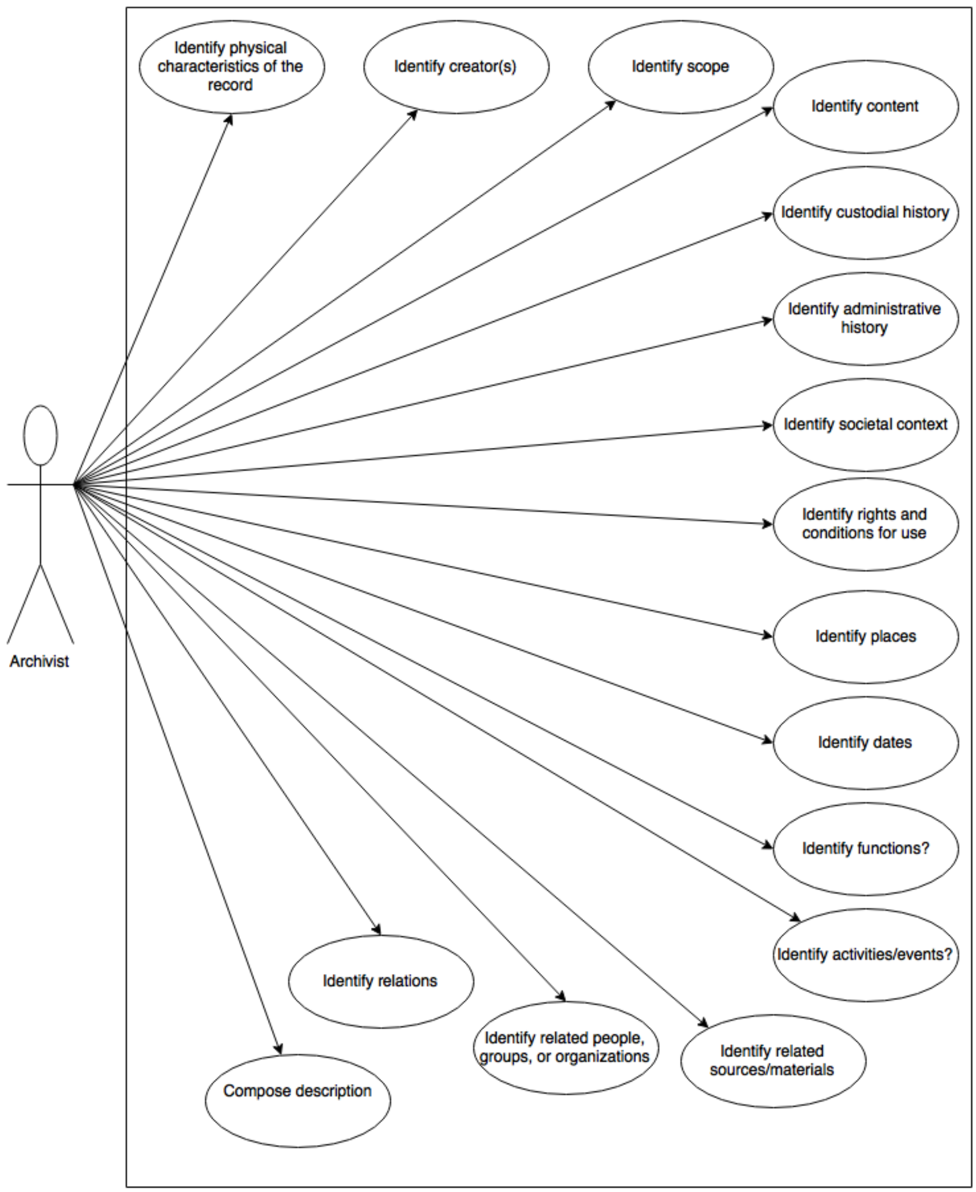
Accession
Accural
AchivementEvent
Appeal
Appoint
Archival Event
Arrangement
Attend
Award
Be-born
Bury
Bussiness
Capture
Communication
Compression
Creation
Custody
Deaccession
Decompression
Decryption
Defend
Deletion
Die
DigitalSignatureValidation
Divorce
Donation
EducationEvent
Employ
End Org
Expedition
Fixity Check
Graduate
hasMilitaryStatus
Heard
Ingestion
LegalEvent
Life Event
Live
Manage Org
Marry
Mention

Merge Org
MessageDigestCalculation
Migration
MilitaryEvent
MilitaryOperation
Normalization
OrganizeBySubject
Partnership
Photograph
ProvenanceUNKnown
Rename Org
Replication
Report
ReportEvent
Resign
Retire
Rule
Start-Org
Teach
Tranaction
Transfer
Transfer-money
Transfer-ownership
Travel
TravelEvent
VirusCheck

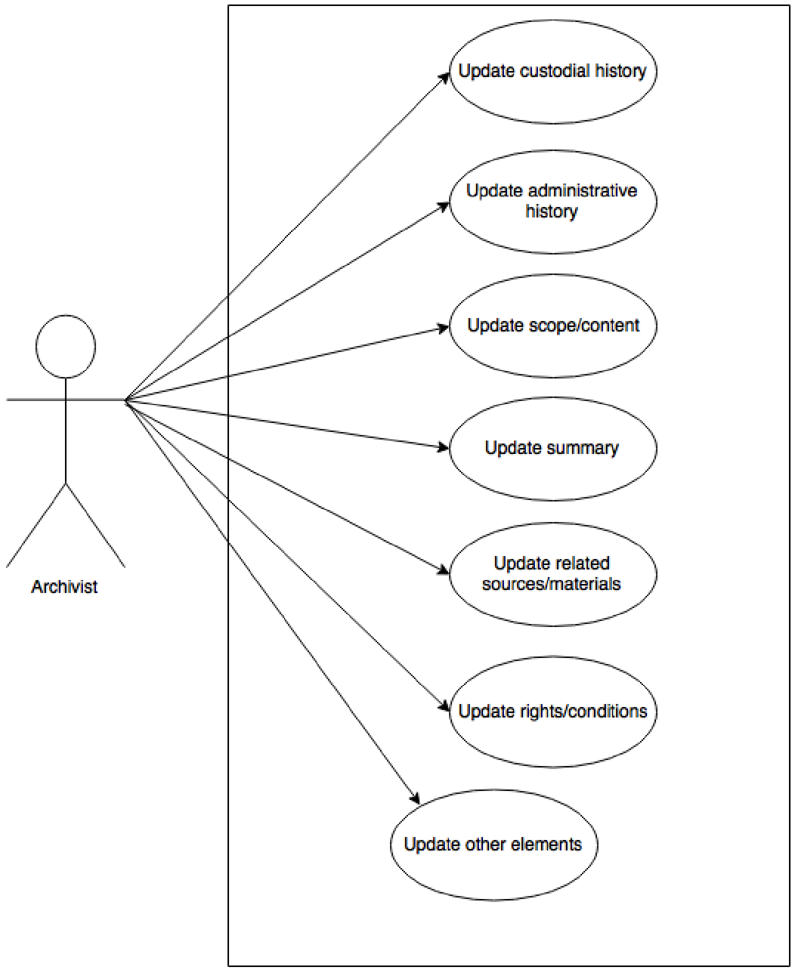
APPENDIX XI.

Use cases

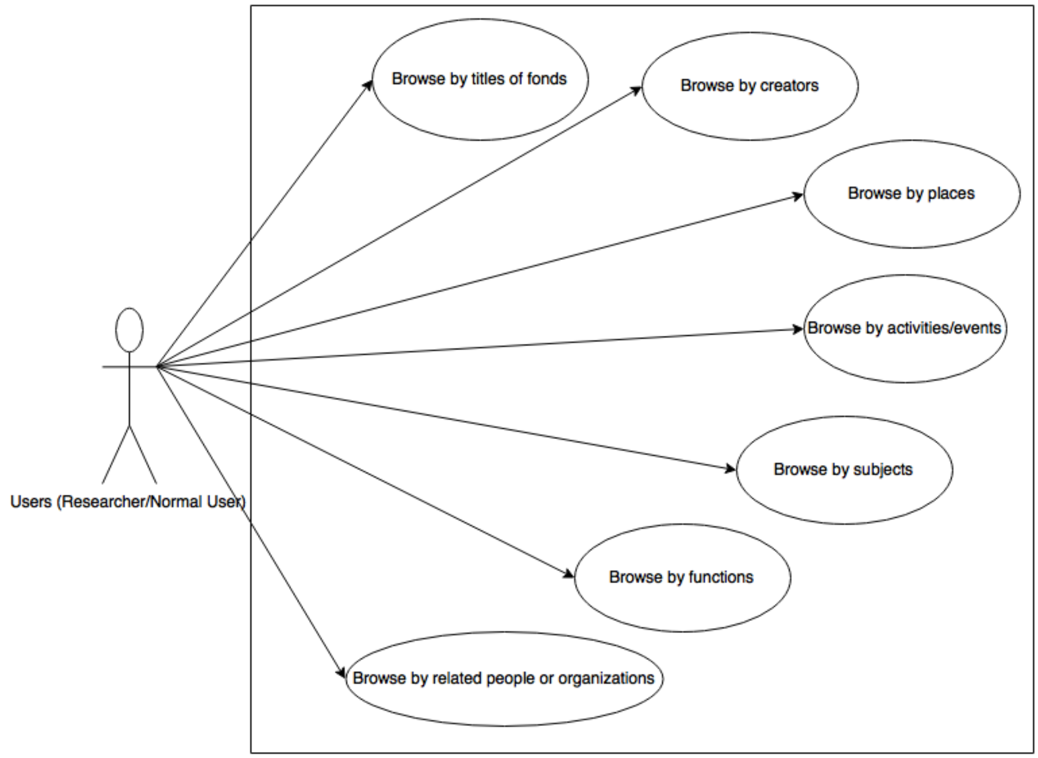
Use case 1: Create Descriptions



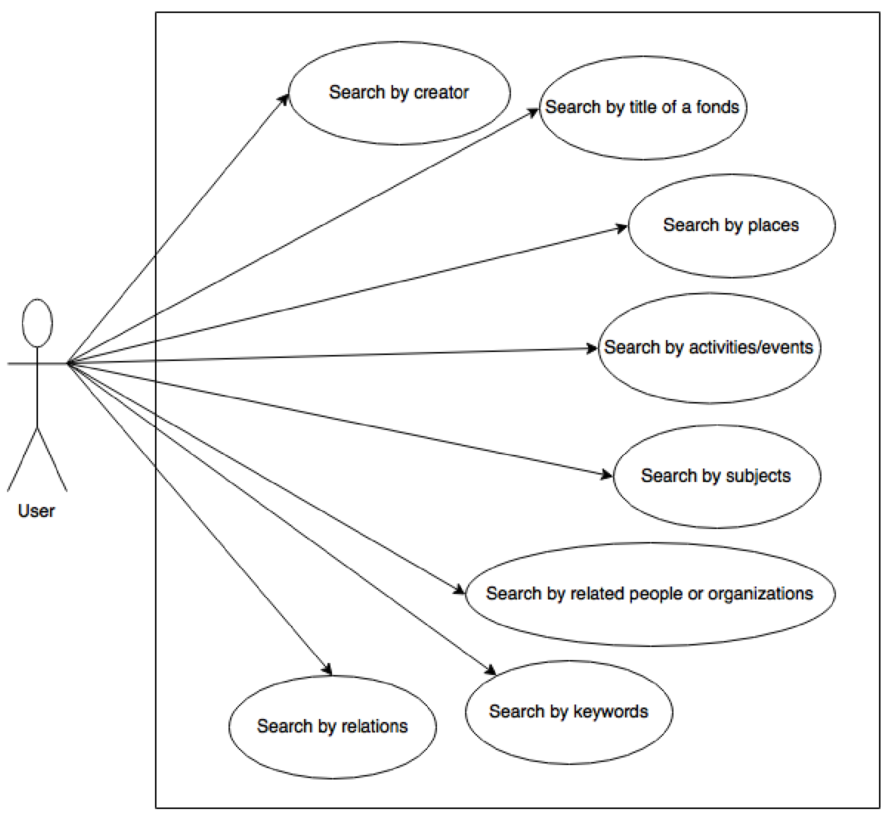
Use case 2: Update archival descriptions



Use case 3: Browse archival descriptions



Use case 4: Search archival descriptions



APPENDIX XII.

Terminology and frequency

Term	Frequency
Corporate body	10
○ Name	4
○ Title	1
○ Other name	2
○ Jurisdiction	1
○ Date of establishing	1
○ Date of enabling	1
○ Date of dissolution	1
○ Resided at	1
○ Has Function	1
○ Internal administrative structure	1
○ Related Person, Corporate body, family	1
Person	10
○ Name	4
○ Title	1
○ Other name	2
○ Gender	1
○ Nationality	1
○ Occupation	2
○ Date of birth	1
○ Date of death	1
○ Has lived at	1
○ Place of Birth	1
○ Place of death	1
○ Religious affiliation	1
○ Political affiliation	1
Family	10
○ Name	4
○ Title	1
○ Gender	1
○ Other name	2
○ Occupation	2
○ Date of birth	1
○ Date of death	1
○ Has lived at	1
○ Place of Birth	1
○ Place of death	1
○ Religious affiliation	1
○ Political affiliation	1

○ Genealogy	2
General Context	1
○ General social context	1
○ General cultural context	1
○ General economic context	1
○ General political context	1
○ General historical context	1
Relation	6
○ Inter-relationship	2
○ Hierarchical relationship	1
○ Temporal relationship	1
○ Family relationship	1
○ Associate relationship	1
History	3
○ Custodial history	1
○ Administrative history	1
○ Biographical history	1
Activity	4
Life Event	2
Role	2
Function	2
Achievement	1
Status	2
Source	1
Record	5
○ Name	1
○ Title	1
○ Date of creation	1
○ Date of modification	1
○ Date of deletion	1
○ Maintenance notes	1
○ Status of records	1
○ Restrictions	1
○ Rights	1
○ Who is involved in processing	1
○ Order	1

APPENDIX XIII.

A glossary of terms

	Description	**Type**
Access	The right, opportunity, or means of finding, using or approaching documents and/or information.	Concept
Accession	The process of formally accepting and recording the receipt of records into archival custody.	Concept (Event)
Accural	An acquisition additional to a unit of description already held by a repository.	Concept (Event)
Achievement	The act of achieving something: the state or condition of having achieved or accomplished something (Merriam-Webster)	Concept
Activity	A series of acts or actions aimed to one purpose (Duranti, Eastwood, & MacNeil, 2002).	Concept (Activity)
Administrative history	To describe the structure and history of the organization which created the records (RAD 1.7B1)	Concept (History)
Affidavits	A type of legal document	Concept (Information Artifact)
Alias	Other form of name	Attribute
Appeal	A legal action	Concept (Event)
Appoint	Assign a job or role to someone	Concept (Event)
Appraisal	The process of identifying materials offered to archives that have sufficient value to be accessioned (Pearce-Moses, 2005).	Concept (Event)
Architecture drawing	A type of information artifact	Concept (Information Artifact)
Archival Description	The process of analyzing, organizing, and recording details about the formal elements of a record or collection of records, such as creator, title, dates, extent, and contents, to facilitate the work's identification, management, and understanding. – 2. The product of such a process. (Pearce-Moses, 2005)	Concept (Information Artifact) Concept (Entity)
Archival Document	See: Record	
Archival Event	Events related processing, managing, and preserving archival materials	Concept (Event)
Archival File	See: File	
Archival Folder	A container may consist of files	Concept (Object Aggregation)

Archival Fonds	See: Fonds	
Archival Function	See: Function	
Archival Record	See: Record	
Archival Role	Role	Concept (Role)
Archival Series	See: Series	
Archives	A place where records selected for permanent preservation are kept; An agency or institution responsible for the function	Concept (Place) Concept (Agency)
Archivist	A professional educated in archival science and/or responsible for the administration of archives. (InterPARES 2)	Concept (Agent/Occupation)
Arrangement	1. The process of organizing materials with respect to their provenance and original order, to protect their context and to achieve physical or intellectual control over the materials. – 2. The organization and sequence of items within a collection (Pearce-Moses, 2005).	Concept (Event)
Associate	Related to …	Relation
Attend	Experience …	Concept (Event)
Author	The person (s), corporate body(ies), or family(ies) chiefly responsible for the intellectual or artistic content of a document. (RAD, 2008)	Concept (Agent with a role)
Authority		Concept (Abstract)
Award	Honor	Concept (Abstract)
BA	A Bachelor of Arts degree	Attribute
Be-born	Birth Event	Concept (Event)
Before_evt	A before relation between two events	Relation
Biographical history	A detailed description of a person's life.	Concept (History)
Brother	A man or boy in relation to other sons and daughters of his parents	Relation
Building	A property	Concept (Property)
Bury	Event	Concept (Event)
Business	Company	Concept (Agent)
Capture	Event	Concept (Event)
Cartographic Plans	A type of information artifact	Concept (Information Artifact)
Case files	A type of information artifact	Concept (Information Artifact)
Cause	Causal relation	Relation
City	A place	Concept
Classification	A scheme	Concept (Abstract)
Collection	An aggregation of records	Concept (Object Aggregation)
Communication	Event	Concept (Event)

Compression	Archival event	Concept (Event)
Copyright	Right	Concept (Abstract)
Co-reference	Two or more things refer to the same thing	Concept
Corporate body	Agent	Concept (Agent)
Correspondence	A type of information artifact	Concept (Information Artifact)
County	Place	Concept (Place)
Creation	A process to create things	Concept (Event)
Creator	Agents who create things in the creation process	Concept (Agent)
Custodial history	A history of custody	Concept (History)
Diary	A type of information artifact	Concept (Information Artifact)
Date of birth	Date of birth	Attribute (Time)
Date of creation	Date of creation	Attribute (Time)
Date of death	The date when a person died.	Attribute (Time)
Date of deletion	Date of deletion	Attribute (Time)
Date of dissolution	Date a corporate body was dissolved.	Attribute (Time)
Date of enabling	The date that certain functions started to work.	Attribute (Time)
Date of establishing	The date when a corporate body has been established.	Attribute (Time)
Date of modification	The date when archival materials have been modified	Attribute (Time)
DaughterOf	Daughter of	Relation
Deaccession	De-accession event	Concept (Event)
Decompression	De-compression event	Concept (Event)
Decryption	Decryption event	Concept (Event)
Defend	Defend in court	Concept (Event)
Degree	Degree	Concept (Abstract)
Deletion	Deletion event	Concept (Event)
Description	A type of information artifact described the objects (representation)	Concept (Information Artifact)
Die	Death Event	Concept (Event)
DigitalSignatureValidation	An event to validate digital signatures	Concept (Event)
Divorce	Leave a marriage relation	Concept (Event)
Document	A type of information artifact	Concept (Information artifact)
Donation	An archival event	Concept (Event)
Donor	Individual who donates records to archives	Concept (Role)
Donor Status	Status of a donor	Concept (Situation)
During_evt	Relation between two events	Relation
Economic Status	Status	Concept (Situation)
Education Event	A specific type of event	Concept (Event)
Electronic record	Records in digital form	Concept (Information artifact)

Employ	Event between a corporate body and a person	Concept (Event)
End Org	An event of dissolution of a corporate body	Concept (Event)
Ephemera	A type of information artifact	Concept (Information artifact)
Essay	A type of information artifact	Concept (Information artifact)
Expedition	A series of translocation event with certain purposes	Concept (Activity)
Family	Family	Concept (Agent)
Features	Specific Attributes	Attribute
Fellowship	A type of Honor	Concept (Abstract)
File	A type of information artifact	Concept (Information artifact)
Finding aids	A type of information artifact	Concept (Information artifact)
Finish_evt	A relation between two events	Relation
Fixity Check	An event to check fixity of digital objects	Concept (Event)
Fonds	The entire body of records of an organization, family, or individual that have been created and accumulated as the result of an organic process reflecting the functions of the creator (2005).	Concept (Aggregation of objects)
Function	The activities of an organization or individual performed to accomplish some mandate or mission (2005)	Concept (Function)
Gender	Male/Female	Attribute
Genealogy	A list or diagram detailing a family tree	Concept (A collection of relations)
General Context	The organizational, functional, and operational circumstances surrounding materials' creation, receipt, storage, or use, and its relationship to other materials; the circumstances that a user may bring to a document that influences that user's understanding of the document. (2005)	Concept (Aggregation of …)
General cultural context	Cultural aspect of general context	Concept (Aggregation of cultural …)
General economic context	Economical aspect of general context	Concept (Aggregation of economic)
General historical context	Historical aspect of general context	Concept (Aggregation of historical stuff)
General political context	Political aspect of general context	Concept (Aggregation of political stuff)
General social context	Social aspect of general context	Concept (Aggregation of social stuff)
Geographical_part	Part of geographical relation	Relation
Graduate	An education event	Concept (Event)
Grant	Honor – monetary	Concept (Abstract)

Group	A group of peoples	Concept (Agent)
HalfBrother	Relation	Relation
Has Function	Indicate corporate bodies have certain functions	Relation
Has lived at	A person has lived at a place	Relation
hasFeature	Has a specific attribute	Relation
hasMilitaryStatus	Has a military status	Relation
hasOccupation	Has an occupation	Relation
hasRole	Has a role	Relation
hasSource	Has a source	Relation
hasStatus	Has a status	Relation
hasSubject	Has a subject	Relation
Health Status	Health condition	Attribute (Situation)
Heard	Event/Legal document	Concept (Event)
Hierarchical relationship	Is-A relations	Relation
History		Concept (History)
Information Artifact		Concept
Ingestion	An event to ingest objects	Concept (Event)
Instant	A type of time	Concept (Time)
Inter-relationship	A type of relation	Relation
Internal administrative structure	A collection of relations related to administrative	Aggregation of relation
Interval	A type of time	Concept (Time)
Judgements	A legal decision	Concept (Information artifact)
Jurisdiction	A region over which an authority may exercise power	Concept (Place)
Land	Real state type of property	Concept (Property)
Legal Documents	A specific type of document to meet legal needs	Concept (Information artifact)
LegalEvent	Event in legal field	Concept (Event)
Letter	A type of information artifact	Concept (Information artifact)
Level of description	The amount of detail or number of elements in a formal description of a work. - 2. The amount of detail in a collection included in a finding aid or catalog record, as determined by the number of hierarchical levels. (2005)	Attribute
Life Event	A specific type of event	Concept (Event)
Live	An event	Concept (Event)
LocatedAt	Located at a place	Relation
MA	Master of Arts degree	Concept (Degree)

Maintenance notes	Notes (a type of information artifact)	Concept (Information artifact)
Manage Org	Event of managing an organization	Concept (Event)
Manuscript	A type of information artifact	Concept (Information artifact)
Map	A type of information artifact	Concept (Information artifact)
Marriage Status	A status of marriage	Concept (Situation)
Marry	An event of getting married	Concept (Event)
Medal	Honor	Concept (Abstract)
Medium	An attribute	Attribute
Meet_evt	A relation between two events	Relation
MemberOf	Membership relation	Relation
Memoradum	A type of information artifact	Concept (Information artifact)
Mention	An event	Concept (Event)
Merge Org	An event of merging organizations	Concept (Event)
MessageDigestCalculation	An event to calculate message digest	Concept (Event)
Migration	An event to migrate from one platform to another	Concept (Event)
Military Status	A status related to military activities	Concept (Situation)
MilitaryOperation	An event related to military activities	Concept (Event)
Minutes	A type of information artifact	Concept (Information artifact)
Money	A special type of property	Concept (Property)
Name	Name	Attribute
Nationality	An attribute of a person's nationality	Attribute
Newspaper clipping	A type information artifact	Concept (Information artifact)
Non-Scholarly Publication	A type of publication	Concept (Publication)
Normalization	An event to normalize objects	Concept (Event)
Occupation	A person's usual or principal work or business, especially as a means of earning a living; vocation	Concept (Occupation)
Organization	A type of corporate body	Concept (Agent)
OrganizeBySubject	A way to organize information artifact	Concept (?)
Origin	The corporate body, administrative unit, family, or individual that creates, receives, or accumulates a body of records, personal papers, or objects (2005).	Concept (Agent)
OriginOrg	A corporate body before changes	Relation
Other name	Other name	Attribute
Overlap_evt	A relation between two events	Relation

Ownership	An ownership relation	Relation
part of	Part-Whole relation	Relation
Partnership	An partnership relation	Relation
Person	1. A human individual. - 2. A legal entity with recognized rights and duties that is treated as a human individual; an artificial person; a legal person; a corporation. (2005)	Concept (Agent)
PhD	Doctor of Philosophy	Concept (Degree)
Photograph	A type of information artifact	Concept (Information artifact)
Photograph Album	A collection of photographs	Aggregation
Physical form	An attributes	Attribute
Place	Place	Concept (Place)
Place of Birth	An attribute of a birth event	Attribute
Place of death	An attribute of a death event	Attribute
Political affiliation	A special attribute related to political aspects	Attribute
Political Status	A status of a person	Concept (Situation)
Postcard	A type of information artifact	Concept (Information artifact)
Preservation	An event to preserve archives	Concept (Event)
Prize	Honor	Concept (Honor)
Processing	An event to process archival records	Concept (Event)
Property	The possession or possessions of a particular owner:	Concept (Property)
Provenance	1. The origin or source of something. - 2. Information regarding the origins, custody, and ownership of an item or collection. (2005)	Concept
Provenance Unknown	Status which provenance is unknown	Concept (Situation)
Publication	A work that expresses some thought in language, signs, or symbols and that is reproduced for distribution (2005)	Concept (Information Artifact)
Record	1. A written or printed work of a legal or official nature that may be used as evidence or proof; a document. - 2. Data or information that has been fixed on some medium; that has content, context, and structure; and that is used as an extension of human memory or to demonstrate accountability. - 3. Data or information in a fixed form that is created or received in the course of individual or institutional activity and set aside (preserved) as evidence of that activity for future reference. (2005)	Concept (Information Artifact)
Related Person, Corporate body, family	Associated Person Associated Family Associated Corporate body	Relation

Relation	Association between two entities	Relation
Religious affiliation	An attribute of an individual	Attribute
Rename Org	A rename event of a corporate body	Concept (Event)
Replication	An archival event to reproduce	Concept (Event)
Report Event	An event	Concept (Event)
Reports	A type of information artifact	Concept (Information artifact)
Resided at	An event (live at)	Concept (Event/Situation)
Resign	Voluntarily leave a job other position	Concept (Event)
Restrictions	Limitations on an individual access to or use of materials (2005).	Concept (Situation)
Retire	An event of a person stops employment completely (wiki).	Concept (Event)
Rights	The legal right of creators to have their works attributed to them and to protect the integrity of their works (2005).	Concept (Rights)
Role		Concept (Role)
Scholarly Publication	A scholarly work that expresses some thought in language, signs, or symbols and that is reproduced for distribution in scholarly media (2005)	Concept (Publication)
Series	A group of similar records that are arranged according to a filing system and that are related as the result of being created, received, or used in the same activity; a file group; a record series. (2005)	Concept (Series)
Silibings	A type of interpersonal relation between brothers and/or sisters	Relation
Social Function	Societal functions	Concept (Function)
Social Role	Societal roles	Concept (Role)
SonOf	An interpersonal relation between a son and a father	Relation
Source	A source publication	Concept (Publication)
Start_evt	A relation between two events	Relation
Start-Org	An event to start an organization	Concept (Event)
Status	Situation	Concept (Status)
Status of records	The status of records	Attribute
Studies	A research investigation	Concept
Sub-fonds	The records of different branches of an organisation or major themes within the papers of an individual (wiki).	Concept (Sub-fonds)
Subject	A principal theme or topic of a work (2005).	Concept (Subject)
Teach	An educational event	Concept (Event)

Temporal relationship	Time related relation	Relation
Title	An attribute	Attribute
Township	A place	Concept (Place)
Transaction	An exchange between two or more entities (individuals or agencies) (2005).	Concept (Event)
Transfer	A transfer event	Concept (Event)
Transfer-money	A transfer event to transfer money	Concept (Event)
Transfer-ownership	A transfer event to transfer ownership	Concept (Event)
Travel	A translocation event	Concept (Event)
Unit of description	A document or aggregation of documents in any physical form, treated as an entity and forming the basis of a single description (2005).	Concept (Information Artifact)
VirusCheck	An event to check for a virus	Concept (Event)
WifeOf	A personal relationship	Relation

APPENDIX XIV.

A complete ad-hoc relation table

Name of Relation	Domain	Range	Inverse Relation
Associate			Associate
At Place		Place	
Before Event	Event	Event	After Event
Brother	Person	Person	Brother
Carried Out At	Event	Place	
Collaborate With	Agent	Agent	
During Event	Event	Event	During Event
Employ	Corporate body	Person	Employ At
Experience	Event	Situation	
Finish Event	Event	Event	Before Event
Graduate From	Person	Educational Organization	
Has Achieved	Agent	Achievement	
Has Ancestor	Person	Person	Has Descendant
Has Attribute (has Feature)			isAttributeOf
Has Cause	Event		isCauseOf
Has Coworker	Person	Person	isCoworkerOf
Has Friend	Person	Person	isFriendOf
Has Function		Function	isFunctionOf
Has Grand Child	Person	Person	isGrandParentOf
Has Half Brother	Person	Person	HasHalfBrother
Has Honor	Person	Honor	
Has Leader	Corporate Body	Person	isLeaderOf
Has lived at	Person	Location	
Has Medical Condition	Person	Situation	isMedicalConditionOf
Has Military Status	Person	Military Status	isMilitaryStatusOf
Has Occupation	Person	Occupational Role	isOccupationOf
Has Participant	Event	Agent	Is Participant Of
Has Partnership	Agent	Agent	hasPartnership
Has Patient	Event	Agent	Is Patient Of
Has Political Affiliation	Person	Political Role	isPoliticalAffiliationOf
Has Professional Partner	Person	Person	Has Professional Partner
Has Professional Role	Person	Professional Role	
Has Project		Project	
Has Property	Agent	Property	

Has Purpose			
Has Religious Affiliation	Person	Religious Role	isReligousAffiliationOf
Has Result	Event		isResultOf
Has Result	Event		isResultOf
Has Role	Agent	Role	isRoleOf
Has Social Role	Agent	Social Role	
Has Son	Person	Person	Has Parent
Has Source		Publication	isSourceOf
Has Spouse	Person	Person	isSpouseOf
Has Status	Event	Situation	isStatusOf
Has Step Brother	Person	Person	Has Step Brother
Has Step Child	Person	Person	Has Step Parent
Has Step Daughter	Person	Person	Has Step Parent
Has Step Father	Person	Person	Has Step Child
Has Step Mother	Person	Person	Has Step Child
Has Step Parent	Person	Person	Has Step Child
Has Step Sibling	Person	Person	Has Step Sibling
Has Step Sister	Person	Person	Has Step Sister
Has Step Son	Person	Person	Has Step Parent
Has Subject		Subject	isSubjectOf
Has Wife	Person	Person	Has Husband
Hierarchical relation (is_a)			isSuperClassOf
In Possession	Agent	Property	In Possession Of
In Relaitonship With	Person	Person	In Relationship
In State (=has Status)			
Involve	Event	Agent	Involve In
Is Member Of	Agent	Family/Corporate body	Has Member
Is Owner Of	Agent	Property	Has Owner
Locate At		Location	
Manifest In	Abstract	Event	
Marry With	Person	Person	Marry With
Meet Event	Event	Event	Meet Event
Occur At	Event	Time	
Overlap Event	Event	Event	Overlap Event
Part Of			Has Part
Place of Birth	BirthEvent	Place	Is Place Of BirthEvent
Place of death	DeathEvent	Place	Is Place Of DeathEvent
Precede	Event	Situation	Follow
Receiving Degree	Person	Degree	
Start Event	Event	Event	End Event

○

APPENDIX XV.

A concept dictionary

Concept Name	Instances	Relation
Collective Artifacts	Finding Aids Archival Document Archival File Archival Folder Archival Fonds Sub-Fonds Archival Record Archival Series Sub-Series Collection Photograph Album	hasMember Associate
Abstract	Achievement Award Fellowship Medal Prize Gift Grant Classification Subject	Associate ManifestIn
Agent	Person Family Corporate Body	Has Role Has Function Involve In Has Status isMemberOf isOwnerOf inPossession hasRelation hasAchieved
Information Artifact	Archival Description Administrative History Description Custodial History Description Architecture Drawing Photograph Minutes Memorandum Manuscript Map Electronic Record Note	part of created by created at time created at place created in context has Attribute Associate

	Maintenance Note Letter Ephemera Essay Dairy Cartographic Plan Correspondence Case File Newspaper Clipping Postcard Report Legal Document Appeal Affidavits Judgement Jurisdiction Publication Non-Scholarly Publication Scholarly Publication	
Certificate	Academic Degree BA MA PhD	givingTo/receiving From
Place	County City Country State Region Township Archives_Place	Part of Associate
Property	Money RealEstate Land Building	isPropertyOf Associate
Role	Archival Role Author Creator Donor Social Role OccupationalRole Professor Archivist Teacher PoliticalRole ReligiousRole Rev.	isRoleOf Associate

	EducationalRole CollegeStudent CharityRole VolunteerRole CommunityRole	
Time	Instant Interval	occuredAtTime
Situation	Restriction Status Marriage Status Health Status Military Status Political Status	
Function		hasFunction Associate
Archival Function		
Government Function	Business Support and Regulation Civil Infrastructure Cultural Affairs Education and Training Employment Environment Finance Management Governance Health Care Immigration Indigenous Affairs International Relations Justice Administration Maritime Services Natural Resources Primary Industries Science Security Sport and Recreation Statistical Services Tourism Trade Transport	hasFunction Part Of Assoicate
Process		
Planned Process	Project Investigation Study	
Creation		Created By Created At Time

		Created At Place Created in Context
Archival Event	Accession Replication Processing Normalization Migration FixityCheck Donation Deletion Accural Appraisal Arrangement Capture Compression Deaccession Decompression Decryption DigitalSignatureValidation Ingestion MessageDigestCalculation Preservation VirusCheck	isEventOf isSubEventOf hasResult precede follow hasStatus
BioEvent = LifeEvent	Appoint Resign Retire Divorce Attend Be-Born Bury EducationEvent Graduate Teach Death EmploymentEvent Employ LegalEvent Heard Live Marry Communication Expedition Mention	isEventOf isSubEventOf hasResult precede follow hasStatus involve hasParticipant hasPatient
MilitaryOperation		
ReportEvent		

TransferEvent	Transfer-Money Transfer-Ownership	
TravelEvent		
ExistEvent		
ExperienceEvent		
GivingEvent		
ReceivingEvent		
BehaveEvent		
OrgChangeEvent	Merge Org Rename Org Manage Org Start Org End Org	isEventOf isSubEventOf hasResult precede follow hasStatus
Feature		isFeatureOf

APPENDIX XVI.

A complete event ontology for archival context

The EOAC ontology can be downloaded from https://archdesc.info/eoac/EOAC.owl.