

# **Searching for Cultural Heritage Information**

## Ontology-based Modeling of User Needs

**Maliheh Farrokhnia**

OsloMet Avhandling 2019 nr 15

**OSLO METROPOLITAN UNIVERSITY**  
STORBYUNIVERSITETET





# **Searching for Cultural Heritage Information**

Ontology-based Modeling of User Needs

Maliheh Farrokhnia



Dissertation for the degree of Philosophiae Doctor (PhD)  
Department of Archivistics, Library and Information Science  
Faculty of Social Sciences  
OsloMet – Oslo Metropolitan University

Spring 2019

CC-BY-SA versjon 4.0

OsloMet Avhandling 2019 nr 15

ISSN 2535-471X (trykket)

ISSN 2535-5414 (online)

ISBN 978-82-8364-164-6 (trykket)

ISBN 978-82-8364-191-2 (online)

OsloMet – storbyuniversitetet

Universitetsbiblioteket

Skriftserien

St. Olavs plass 4,

0130 Oslo,

Telefon (47) 64 84 90 00

Postadresse:

Postboks 4, St. Olavs plass

0130 Oslo

Trykket hos Byråservice

Trykket på Scandia 2000 white, 80 gram på materiesider/200 gram på coveret

## **Acknowledgments**

It is my great pleasure to thank the many people who have contributed in one way or another to my success in completing my doctoral dissertation and degree. First and foremost, I would like to express my deepest gratitude and appreciation to my supervisors, Professor Nils Pharo and Professor Trond Aalberg. My research has benefited from their invaluable expertise, support, and encouragement.

Special thanks to Professor Rahmatollah Fattahi for kindly reviewing and providing feedback on my research proposal, which was one of the main determining factors for my acceptance into the PhD programme here at OsloMet. I also want mention Iran Software & Hardware Co. (NOSA), from which I gained much inspiration for doing my doctoral research.

Thank you to Dr. Christian Emil Ore, Dr. Koraljka Golub, and Dr. Michael Preminger for agreeing to serve as members of my PhD committee and for their thorough and insightful review of my dissertation. I would also like to thank Professor Pia Borlund for serving as the opponent for my final seminar and challenging me and helping me gain new insights and understandings. My gratitude also goes to the Department of Archivistcs, Library and Information Science (ABI), the former and current Head of Department, Liv Gjestrum and Tor Arne Dahl, and to the former and current PhD Programme Coordinator, Professor Ragnar Andreas Audunson and Professor Katriina Byström.

I am especially grateful for the grant to study abroad from the Faculty of Social Sciences that allowed me to spend over six months as a visiting scholar at the School of Information and Library Science (SILS) at the University of North Carolina at Chapel Hill. I sincerely appreciate SILS for hosting me, especially Professor Dian Kelly for welcoming me into her LAB, Professor Barbara Wildemuth for her mentoring, and Kaitlyn Murphy for making my stay memorable.

I was lucky to have a fabulous bunch of fellow PhD candidates. Thank you for the enjoyable companionship, stimulating discussions, and support. I would like to thank all the friends that I have made since the start of my doctoral studies. Thank you for all the great times that we have shared! This experience would not have been the same without you. I also want to express deep gratitude to my dear friends, Jamie Johnston and Debbie Maron who kindly proofread parts of this dissertation. I would also like to thank Martin Helsø for his help with the LATEX program.

Additionally, I would like to express my gratitude to my wonderful parents and dear brothers, Ali and Ehsan. My deepest gratitude goes to my lovely sister, Mina, for her moral and emotional support, especially during the health-related challenges that I faced during my doctoral studies. I would not be

where I am today without all of their continuous love, support and encouragement.

Finally, and above all, I wish to acknowledge the trust and generosity of the participants who were willing to take part in this study.

## **Abstract (English)**

This PhD dissertation investigates users' actual information needs when they search for cultural heritage information and examines how their needs map to ontologies such as CIDOC Conceptual Reference Model (CRM).

With the invention of the semantic web and increasing demand for global search across cultural heritage data, ontologies such as CIDOC CRM have been developed to provide an extensible semantic framework to formally describe the implicit and explicit relationships of cultural-heritage concepts. This ontology provides a basis for meaningful integrated access to semantically heterogeneous data. Currently, ontologies are developed by domain experts and system developers. In this project, qualitative research incorporating user's perspective and actual needs is used to complement our understanding of how well the ontologies model user's needs.

With a naturalistic approach, this study has conducted contextual inquiry to explore user's information needs. Analyzing the qualitative data reveals that the information needs can be grouped into four emergent categories of Fact-finding, Entity-finding, Resource-finding, and Context-finding. As the result of analysis, the model of semantic information needs has been introduced that can be considered as an extension to Taylor's levels of information needs.

Based on the ontological analysis of information needs, some general patterns have been discovered and projected onto CIDOC CRM.

The findings of this research show that ontology-based information systems can be built upon a limited types of entities and properties to meet a broad range of information needs.

**Keywords:** Information need, CIDOC CRM, Modeling

## Sammendrag (Norsk)

Denne PhD-avhandlingen undersøker brukernes informasjonsbehov når de søker etter kulturarvsinformasjon og hvordan disse kan mappes til ontologier som CIDOC Conceptual Reference Model (CRM).

Utviklingen av den semantiske weben og økende behov for å foreta globale søk i kulturarvdata har ført til utviklingen av ontologier som CIDOC CRM. Disse utgjør utvidbare semantiske rammeverk som kan beskrive formelt de implisitte og eksplisitte relasjonene mellom kulturarvbegreper. Denne ontologien er basis for meningsfull integrert tilgang til semantisk heterogene data. Ontologier utvikles av domene-eksperter og systemutviklere. Dette prosjektet undersøker, ved hjelp av kvalitativ forskning på brukeres faktiske informasjonsbehov, hvor godt ontologier modellerer sluttbrukeres behov.

Forskningsprosjektet har, med et naturalistisk utgangspunkt, undersøkt informasjonsbehovene. Analyse av de kvalitative dataene avslører at informasjonsbehovene kan deles inn i fire kategorier: fakta-finning, enhets-finning, ressurs-finning og kontekst-finning. En modell (Model of semantic information needs) har blitt utviklet, basert på dataanalysen. Modellen er en utvidelse av Taylors nivåer av informasjonsbehov.

Ved å foreta en ontologibasert analyse av informasjonsbehov har noen generelle mønstre blitt avdekket og projisert på CIDOC CRM.

Funnene i dette prosjektet viser at ontologibaserte informasjonssystemer kan bygge på et begrenset sett av enheter og egenskaper for å imøtekomme et bredt spekter av informasjonsbehov.

# Contents

<b>Contents</b>	<b>1</b>
<b>List of Tables</b>	<b>5</b>
<b>List of Figures</b>	<b>7</b>
<b>1 Introduction</b>	<b>9</b>
1.1 Problem statement and motivation . . . . .	10
1.2 Research questions and objectives . . . . .	11
1.3 Approach . . . . .	11
1.4 Thesis structure . . . . .	12
<b>2 Theoretical background</b>	<b>15</b>
2.1 Information needs . . . . .	16
2.1.1 Traditional system-centered approach . . . . .	16
2.1.2 User-centered cognitive approach . . . . .	17
2.1.3 Motivation of information needs . . . . .	24
2.1.4 Types of information needs . . . . .	27
2.2 Information organization . . . . .	32
2.2.1 Cultural heritage information and increasing trend to be integratedly accessible . . . . .	32
2.2.2 Semantic-network-based systems . . . . .	33
2.2.3 Ontology vs. semantic network . . . . .	34
2.2.4 Organizing strategies and ontologies . . . . .	34
2.2.5 Ontologies' compatibility with users' mental models . . . . .	40
<b>3 Methodology</b>	<b>43</b>
3.1 Naturalistic Inquiry . . . . .	44
3.2 Study's design . . . . .	45
3.2.1 Contextual Inquiry . . . . .	45
3.2.2 Sampling . . . . .	47
3.2.3 Instrument . . . . .	49

## 2 Contents

---

3.3	Pilot study . . . . .	49
3.4	Data Collection . . . . .	50
3.5	Data Analysis . . . . .	51
3.5.1	Phase 1: Thematic analysis and open coding . . . . .	53
3.5.2	Phase 2: Modeling the progression of information needs . . . . .	53
3.5.3	Phase 3: Creating the stories . . . . .	57
3.5.4	Phase 4: Ontological analysis . . . . .	59
3.5.5	Phase 5: Final analysis . . . . .	60
3.6	Limitations . . . . .	61
3.7	Validity and reliability . . . . .	61
3.8	Researcher position . . . . .	62
3.9	Ethics approval . . . . .	62
<b>4</b>	<b>Results</b>	<b>63</b>
4.1	Story 1: User is interested to know about a particular fashion designer . . . . .	64
4.1.1	Introduction . . . . .	64
4.1.2	Query 1.1: <i>“Intellectual property fashion”</i> . . . . .	67
4.1.3	Query 1.2: <i>History of property/ histoire de la propriété intellectual</i> . . . . .	68
4.1.4	Query 1.3: <i>Elsa Schiaparelli</i> . . . . .	68
4.1.5	Query 1.4: <i>Schiaparelli</i> . . . . .	73
4.1.6	Query 1.5: <i>Elsa Schiaparelli</i> . . . . .	74
4.1.7	Query 1.6: <i>Christian Bérard</i> . . . . .	75
4.1.8	Query 1.7: <i>Christian Bérard</i> . . . . .	75
4.2	Story 2: User would like to find information about a particular law . . . . .	76
4.2.1	Introduction . . . . .	76
4.2.2	Query 2.1: <i>Rural employment guarantee</i> . . . . .	78
4.2.3	Query 2.2: <i>“human right” and ecology</i> . . . . .	79
4.3	Story 3: User is looking for information about social system . . . . .	80
4.3.1	Introduction . . . . .	80
4.3.2	Query 3.1: <i>Evolution Social System</i> . . . . .	81
4.3.3	Query 3.2: <i>Tom R. Burns</i> . . . . .	82
4.3.4	Query 3.3: <i>Tom R. Burns</i> . . . . .	83
4.4	Story 4: User is interested in information about translating of books . . . . .	83
4.4.1	Introduction . . . . .	83
4.4.2	Query 4.1: <i>Actor Network Theory</i> . . . . .	84
4.4.3	Query 4.2: <i>“Jennifer Egan Bakke”</i> . . . . .	85
4.4.4	Query 4.3: <i>“David Vann Jord”</i> . . . . .	85
4.5	Story 5: User is interested in information about a particular travel writer . . . . .	86
4.5.1	Introduction . . . . .	86
4.5.2	Query 5.1: <i>Travel writing and Canad* and North</i> . . . . .	88

---

4.5.3	Query 5.2: <i>Richard Harrington</i> . . . . .	89
4.5.4	Query 5.3: <i>Richard Harrington</i> . . . . .	90
4.5.5	Query 5.4: “ <i>Richard Harrington</i> ” AND “ <i>face of the arctic</i> ” . . . . .	92
4.6	Story 6: User wants to learn about value chain . . . . .	93
4.6.1	Introduction . . . . .	93
4.6.2	Query 6.1: <i>value chain</i> . . . . .	93
4.6.3	Query 6.2: <i>value chain</i> . . . . .	93
4.6.4	Query 6.3: <i>value chain</i> . . . . .	94
4.7	Story 7: User wants to know about gender and rights to the city . . . . .	94
4.7.1	Introduction . . . . .	94
4.7.2	Query 7.1: <i>Gender and Power Connell</i> . . . . .	95
4.7.3	Query 7.2: <i>women right to the city</i> . . . . .	96
4.8	Story 8: User is interested in archaeology . . . . .	96
4.8.1	Introduction . . . . .	96
4.8.2	Query 8.1: <i>materiality archaeology</i> . . . . .	97
4.9	Story 9: User is looking for information about development of children’s libraries . . . . .	98
4.9.1	Introduction . . . . .	98
4.9.2	Query 9.1: <i>sociology AND childhood AND Giddens</i> . . . . .	99
4.9.3	Information needs sought manually . . . . .	100
4.10	Conclusion . . . . .	101
<b>5</b>	<b>Analysis and Discussion</b> . . . . .	<b>103</b>
5.1	Holistic perspective of users’ information needs . . . . .	104
5.2	Types of information needs . . . . .	109
5.2.1	Fact-Finding . . . . .	110
5.2.2	Resource-Finding . . . . .	111
5.2.3	Entity-Finding . . . . .	113
5.2.4	Context-Finding . . . . .	114
5.3	General patterns of users’ information needs . . . . .	116
5.3.1	Category 1: Three principal notions of actor, thing, or event as the target of information need . . . . .	117
5.3.2	Category 2: Related information on actor, thing, or event . . . . .	129
5.4	Model of Semantic Information Needs . . . . .	133
<b>6</b>	<b>Conclusion</b> . . . . .	<b>137</b>
6.1	Contributions . . . . .	139
6.1.1	Contributions to theory . . . . .	139
6.1.2	Methodological contributions . . . . .	140
6.1.3	Contributions to practice . . . . .	141
6.2	Future research . . . . .	142

## 4 Contents

---

<b>References</b>	<b>145</b>
<b>Appendix A Interview guide</b>	<b>159</b>
<b>Appendix B Consent Form</b>	<b>161</b>
<b>Appendix C Approval from Norwegian Center for Research Data, NSD</b>	<b>165</b>

# List of Tables

3.1	Profile of the participants . . . . .	48
5.1	A compact summary of participant's information needs . . . . .	105



# List of Figures

2.1	Information need type continuum . . . . .	28
3.1	Holistic Model of Information-Needs Progression . . . . .	54
3.2	Sample of the Information Needs Progression Modeling . . . . .	57
3.3	Sample of ontological analysis of an information need . . . . .	60
5.1	Cultural heritage information needs of users . . . . .	110
5.2	Network of persons working on a concept . . . . .	119
5.3	Network of collaborators on an activity . . . . .	120
5.4	Person's collaborators . . . . .	121
5.5	Addressee of a person's correspondences . . . . .	122
5.6	Works created by a person . . . . .	123
5.7	Works inspired by information object . . . . .	124
5.8	Discussions around a person and/ or person's created work . . . . .	125
5.9	Discussions around a concept . . . . .	126
5.10	Discussions around an activity . . . . .	127
5.11	Translation of a work . . . . .	128
5.12	Events and activities related to a person . . . . .	129
5.13	Information related to physical man-made thing . . . . .	131
5.14	Information related to a person . . . . .	132
5.15	Model of Semantic Information Needs . . . . .	134



Chapter 1

# **Introduction**

### 1.1 Problem statement and motivation

With the invention of the semantic web, the interest in making publicly available the tremendous wealth of information kept in archives, libraries, and museums (ALM), i.e., “memory institutions”, has been growing. Moreover, any piece of information does not stand alone in the information world. Each is loaded with history, authorship, and creation context in relation to other pieces of information scattered across different memory institutions. As more and more varied cultural-heritage information (i.e., information recorded through various information resources preserved in memory institutions) becomes available on the web, the borders of domains, information carriers, and information providers are blurring. Consequently, the demand for global search and data integration across heterogeneous sources of cultural-heritage information is increasing (Doerr, 2005). To facilitate the integration of this data, ontologies such as CIDOC Conceptual Reference Model (CRM) have been developed. CIDOC CRM provides an extensible ontology for concepts and information found in cultural-heritage data in different domains. Through such ontology-based information systems, memory institutions can describe cultural data with interoperable semantics, and this semantic interoperability potentially can provide meaningful integrated access to cultural-heritage information. Currently ontologies have been developed by domain experts and system developers. Although Oldman (2014) states that ontologies such as CIDOC CRM have been developed as a result of empirical analysis of existing models of practice that, themselves, have been developed over a considerable period of time, ontologies remain far from users’ perspective and should not be made visible to end users (Le Boeuf, 2013). Ontologies are well-developed to manage highly granular data in information systems’ lower layers, and the huge complexity of such granularity is not understandable to the end user. For example, users never will be able to deal with numerous classes and properties in an ontology like CIDOC CRM. We need a transformation between layers of ontology-based information systems so that users can make more sense of them and use them in these systems’ application layers. To contribute to developing such a transformation, we need to understand what users actually need and how much of this complex semantic data provided in the lower layers of ontology-based information systems should be accessible to users. Understanding users’ information needs that underlie a query long has been recognized as key to effective information retrieval. However, Ingwersen and Järvelin (2005) criticize Information Seeking studies as they rarely include elements of information system design, which prevents the results of those studies from being applicable to information systems. They suggest a need to extend Information Seeking studies toward information systems and their designs. In such a new global context, we need to refocus strongly on users’ needs and expectations when developing solutions for making semantic information accessible. Ontological analysis of users’ information needs has been explored, but not very systematically, through empirical research, with little extant research available (e.g. Henniscke, 2017). We do not know enough about how well users’ information needs can be mapped into ontologies, and we need empirical research to realize how users’ cognitive structures correspond to ontology developers’ cognitive structures.

This thesis aims to investigate users' actual information needs in more depth, and through ontological analysis explore common patterns of information needs that can be used in ontology-based information systems' application layers to provide more effective information retrieval.

## 1.2 Research questions and objectives

In conducting this research, the following questions are addressed:

1. How do users search for cultural-heritage information?
2. What are users' expectations across information systems, and what types of information do they seek?
3. What common patterns can be discerned within users' information needs?
4. How can users' information-need patterns be projected onto CIDOC CRM and its domain extensions?

Following a naturalistic approach, through conducting contextual inquiries and ontological analysis, the specific objectives of this thesis are:

- To determine whether any similarities exist between ontology structures and researchers' information needs by exploring users' conceptualization of cultural-heritage information and their actual information needs.
- To identify characteristics and patterns in users' needs through ontological analysis of their actual information needs.
- To see whether and how well users' information needs can be projected onto ontologies such as CIDOC CRM and its domain-specific extensions.
- To model users' information needs through ontological representations that can be used in the application layers of future ontology-based information systems.

## 1.3 Approach

Exploring possible approaches to address this study's objectives, we have chosen the following concepts on which to build our work:

- **Holistic approach toward users' information needs**

Users' information needs have been viewed independent of any specific information-carrier type, specific domain or field, or specific information system. Moreover, users' information

needs have been analyzed not as one single phenomenon, but in the context of information needs while considering all related information needs that are expressed either through some search tasks or explained by the user.

The methodological approach to this research goes beyond the compromised level of information needs (i.e., what users express as their queries in information systems). By conducting a qualitative study using contextual inquiry, the present study aims to discover users' real information needs, expressed as queries, and to elicit their other related information needs based on their corresponding tasks, interest, and curiosity.

- **Ontology-based approach toward understanding users' information needs**

This study describes users' information needs in terms of concepts (entities) and relationships (properties) defined by a formal ontology. For ontological analysis of users' information, we have chosen CIDOC CRM as a general ontology and its developing domain-specific extensions. This approach helps in understanding and presenting users' information needs in a semantic way that can be useful for more effective information retrieval in ontology-based information systems.

## 1.4 Thesis structure

The thesis is structured as follows:

Chapter 2 provides the theoretical background used as the basis for the present study in two parts: information needs and information organization. The first part starts by describing the two approaches to the concept of information needs in information retrieval (IR) studies, including the traditional system-centered approach and the user-centered approach. It then continues with a discussion on what triggers and motivates information needs. Finally, it provides a review of different types of information needs introduced in extant literature. The second part is dedicated to the organization of information and different kinds of models and ontologies developed with the invention of the semantic web to organize cultural-heritage information.

Chapter 3 examines the methodology used in the present study. It provides an overview of the naturalistic approach vs. the positivistic approach, and also compares different methods used in information-need studies. It then provides a detailed description of study design, the qualitative method of contextual inquiry used to collect data, sampling process, and data collection. Different phases of data analysis are described, including thematic analysis and open coding, modelling the progression of information needs- in which we also introduce a particular tool (holistic model of information-needs progression) that was developed to analyze the contextual interviews-, generating participants' information-needs stories, ontological analysis, and a final analysis, including a comparative analysis.

Chapter 4 presents findings and results in the form of narrative stories that impart contextual information to better understand participants' expressed information needs. The stories provide

detailed descriptions of different information needs that participants expressed during the contextual-inquiry session. Each story starts with an introduction, including participants' research background, work task, and search behaviors. The story then continues by describing different search episodes and related information needs. The chapter also discusses the present study's limitations, as well as its validity and reliability.

Chapter 5 discusses the results of the analyses, with respect to extant literature, in four parts. Through a holistic view, participants' information needs are discussed as wholes in which different interconnected information needs are understood in three parts – queries, primary needs, and secondary needs. Four emergent types of information needs – fact-finding, resource-finding, entity-finding, and context-finding – are introduced in this chapter. The chapter then provides an ontology-based presentation of general patterns of information needs that emerged in this study and concludes by introducing the model of semantic information need, which is an extension to Taylor's levels of information need.

Chapter 6 summarizes the study's principal points and its contributions in three parts – theory, methodology, and practice. The chapter concludes the dissertation with some suggestions for future research.



## Chapter 2

# **Theoretical background**

Developing effective information systems requires a solid understanding of how users' conceptualize their information needs and what expectations users' have of the information systems, as well as how information is best organized within the systems to meet users' needs.

This chapter discusses these two aspects of developing effective information systems: users' information needs and information organization. The first part is dedicated to the concept of information needs, in which the two approaches to information retrieval, the concept of relevance, information needs' motivations, and types of information needs are discussed. The second part of this chapter deals with the concept of information organization and discusses different issues, including cultural-heritage information and an increasing trend to be integratedly accessible, semantic-network-based systems and ontologies vs. semantic network, organizing strategies and ontologies, and compatibility of ontologies with users' mental models.

## 2.1 Information needs

Since the late 1950s, researchers have worked to understand how to best design computerized systems to make information accessible to meet users' information needs. They studied information and information retrieval (IR) through two different approaches: the traditional system-centered approach and the user-centered cognitive approach discussed in this paper.

### 2.1.1 Traditional system-centered approach

For a while after the computerization of information systems, most studies were considerably constrained by the system's definition of needs. In the system-centered approach, according to the study of Dervin and Nilan (1986, p. 17), information needs are defined "[not] as what users think they need but rather in terms that designate what it is in information system that is needed." In other words, the focus of this approach is on what the systems possessed rather than on what users are need or the gaps in knowledge or understanding that they face (Dervin & Nilan 1986).

Information in a system-centered approach, according to Shannon and Weaver (1949), is viewed as something objective that exists outside the user. Information is treated as something external to be discovered and retrieved from information systems and presented to the user to meet his/her internal needs, preferably at the right time and in the right format to reduce user's ambiguity (Morris 1994). The goal of information seeking and retrieval through the traditional paradigm of system-centered, as Morris (1994) states, is to "find the 'external information reality' that corresponds to our 'internal need'". One of the best known examples of research on system-centered approaches is the Cranfield studies in the 1960s that assumed that the vast majority of users require a few relevant materials, in which content experts determine relevance, no matter who the user is (Cleverdon 1991, Katz 2013). Kuhlthau (1991, p. 361) also criticizes information retrieval in the system-centered approach by arguing that retrieval in this approach is focused on "questions that best match the system's representation of texts rather than responding to user's problems". The major critique to this traditional information retrieval is

the difference between the nature of the content of information system and information needs of users regarding ambiguity and order. As Kuhlthau (1991, p. 361) explains, “bibliographic paradigm is based on certainty and order, whereas users’ problems and needs are characterized by uncertainty and confusion”. Dervin (1981) identifies the underlying assumptions the system-centered understanding of information needs implies: It is assumed that messages can have a direct impact, that they can ‘get into’ receivers the same way they left the sources and that they impact the receivers all in the same way. It assumes that a message is received the same way by the source and receiver and similarly by all receivers. It assumes that there is nothing unique about the receiver that will impact his or her use of the message. It assumes that there is no cognitive process intervening between message and use. Behind all these assumptions, is the core assumption dealing with the nature of information, that information can be dropped into people’s heads as if people’s heads were empty buckets. He concludes that to make this assumption “it must be assumed that information is a thing rather than a construction, that it exists independently of observers and has an inherent, correct, absolute, and isomorphic relationship to the reality it describes” (Dervin 1981, p. 74).

### 2.1.2 User-centered cognitive approach

The user-oriented understanding of the IR paradigm was developed as an alternative to the system-oriented IR approach. In the user-centered approach, information is subjective, situational, holistic and cognitive, and should be understood within specific contexts (Dervin & Nilan 1986). According to Dervin (1992, p. 81), information is not something that has constant meaning, describes a given reality in an absolute way, or can be transmitted from the information producer to the information receiver through channels; “Rather, information is constructed”. Although, the information may look similar, the way different users understand, experience and make sense of it is different for each individual and in different times and spaces. According to Kelly (1969)’s Personal Construct Theory, people actively develop personal constructs as they are confronted with new events and as they attempt to make sense of the world. Different people make sense and interpret the same events in their own way. Personal construct theory suggests that “human behaviour is to be understood in a context of relevance” (Kelly 1969, p. 12).

In user-centered approach, the focus on document representation and associated search techniques shifts to understanding of users and their problems (information needs), their beliefs and individual differences, and their knowledge structures and cognitive process in information seeking situations (Ingwersen 1982, Brookes 1980, Borgman 1985, Belkin 1978, Belkin et al. 1982, Belkin 1990, Dervin 1983a;b). Attention to cognitive processes in information retrieval dates back to 1960s in the early work of Taylor (1967) and Paisley and Parker (1965), where the importance of understanding users’ knowledge structures and their interactions during the information retrieval process was recognized. Edwin Parker and William Paisley, according to Ruthven & Kelly (2011, p. xxvii), were among the pioneers in the mid 1960s to introduce an interdisciplinary approach including psychology, communication and information science to a

user-centered information retrieval study. Here, some cognitive perspectives will be discussed. However, the discussion is not exhaustive, but representative. I particularly tend to limit the discussion to contributions that have inspired the current study.

**Taylor** Taylor's work serves as a theoretical foundation for several user-oriented studies in information seeking and retrieval. According to Hjørland (1997), when Taylor's hypothesis first was published, it was a great inspiration because it provided a new methodological approach by combining information-science problems with psychology. In his article "The process of asking questions", published in the *Journal of the American Society for Information Science (JASIS)*, Taylor (1962) identified four levels of information needs (question formation): 1) visceral; 2) conscious; 3) formalized; and 4) compromised. He believes that information needs develop through all four levels, from "visceral need" to "conscious need" to "formalized need" to "compromised need" before the user formally encounters an information system. These four levels overlap with each other along the question spectrum. The first level is the actual, but unexpressed need, which Taylor calls the visceral need. It is the conscious, or even unconscious, need for information in which the user recognizes some deficiency in his or her knowledge structure, but may not be able to define this need cognitively. The information need in this level is not a question yet, but "only a vague sort of dissatisfaction" that is probably not linguistically expressible (Taylor 1968, p. 182). The second level is the conscious need, in which the user has a cognitive grasp of the need and can characterize the deficiency. The information need can be an ambiguous statement in this level. The third level is the formalized need, in which the user can provide a rational, coherent, and formal description of the need in concrete terms. At this level, the user can articulate a clear statement of his or her information need and "may or may not be thinking within the context or constraints of the system from which he wants information", as this level comes before formally encountering an information system (Taylor 1968, p. 182). The fourth level is the compromised need, in which the articulated need is entered into an information system as a query. Considering the information system's constraints and the available information, at this level, the user recasts his or her question in anticipation of what the information system can deliver.

In his Question-Negotiation Framework (QNF), Taylor (1968) explains that a question or information need develops in the user's mind and cycles through these four general levels of need, and that the information system must negotiate the user's question from the compromised expression of the need, through the formalized need, and possibly even to the conscious need. The first level of information need, visceral need, is traceable, but probably not easily reachable. In the question-negotiation process, the user tries to describe not something that he or she knows, but rather something that he or she does not know (Taylor 1968, p. 180). Thus, to realize the user's actual question through question negotiation, as Taylor notes, it is important to make five determinations: 1. the subject of the user's interest, the subject's extent, and its limits and structure to be able to define, expand, and narrow the question; 2. the user's objective and motivation, which provide the opportunity to determine the user's perspective and characteristics

of possible answers to the question (according to Taylor, users cannot always define what they want but they can discuss why they want it); 3. the user's personal characteristics and background, which can provide some contextual data; 4. the relationship between the question's description and information organization within the system, which helps restructure the query to fit the information and its organization, structure, and associations, as it is organized within the information system to be found via an efficient search; and 5. the user's anticipated answers. The user has some picture in mind about the type of answer sought regarding its format, data, size, etc. (Taylor 1962). However, this mental picture can be altered, and the user may change his or her question as he or she becomes aware of the information system's capabilities. Search strategies also might change, and he or she might need to place limits on his or her question (Taylor 1968).

Cole (2015, p. 4117) calls Taylor's four levels of information needs "the apogee of the need concept". The explicit and implicit impact of Taylor's information needs and information-negotiation model can be viewed in other researchers' extant studies (such as Markey (1981), Belkin et al. (1982), Dervin and Dewdney (1986), Kuhlthau (1988; 1991), Gross (1995), and Cole (2011)). For example, extending Taylor's model of information need, Markey (1981) proposes a model of negotiated and isolated states of the aforementioned Taylor model's four levels. Her proposed model accommodates bearers of needs, including the librarian and user, and the interaction between them. She supplements Taylor's model by making distinctions on needs as either isolated or negotiated. Her model's initial level refers to the visceral need as an isolated level in which no interaction happens, with the user being the only bearer of the visceral need. Her model then assumes that the user goes further to the next levels, i.e., the isolated conscious level and maybe the isolated formalized level, before encountering the librarian. After negotiation with the librarian, the remainder of the negotiation takes place as a collaborative effort, and the ensuing levels are termed negotiated, rather than isolated.

In their discussion of the Anomalous State of Knowledge (ASK), Belkin, Oddy, and Brooks (1982) explain ASK as an extension of Taylor's idea of question negotiation. Addressing the question-negotiation proposed by Taylor as being one of the most important tasks of a librarian, Dervin and Dewdney (1986) introduce "neutral questioning" as a communication technique for interviewing library users. Neutral questions are open in form, and they help the librarian learn about "the nature of the underlying situation, the gaps faced, and the expected uses, i.e., the three elements of the Sense-Making model" (Dervin & Dewdney 1986, p. 4). Kuhlthau (1988; 1991), inspired by Taylor's model and other information-seeking models, proposes her six-stage Information Search Process Model (that will be explained in detail later). Gross (1995) uses Taylor's levels of information need to develop her imposed-query model. The imposed query is one in which an imposer develops a query, which is given to an agent to transact or resolve. The imposed query, as Gross (1995, p. 237) describes, "is negotiated and transacted outside the purview of the person originating it". The model illustrates the process of the query being transferred from the imposer to the agent, transacted, then returned to the imposer. However,

Taylor's work really had not been challenged and tested (Markey 1981, Hjørland 1997) until the 1990s, when Hjørland (1997) explicitly challenged Taylor's work on information need development.

Hjørland (1997, p. 163) disagrees that four levels of question formation, according to Taylor's hypothesis, develop along a continuum and argues that "it is hardly possible that the information need, as Taylor implies, can progress in a continuous manner, since a given piece of information may disturb the whole problem that triggered the need". He also believes that it is not the information need that develops inside the user's mind, but rather knowledge of the problem area that triggers the information need. He even proposes that it is "cognitive development" more than "information-need development" in the user's mind. In a user study using Taylor's information-need levels as the framework, Pettigrew (2000) characterizes information interactions between nurses and seniors at foot clinics, with her findings indicating that nurses can recognize different types of needs, including seniors' visceral and conscious needs. To support her claim, she explains that nurses "pick up on needs that the seniors themselves may not realize they have and also on needs that the seniors are unable to express in concrete terms" (Pettigrew 2000, p. 67). Cole (2011) analyzed Taylor's four-level information-need model both horizontally and vertically. In his horizontal interpretation, or phase approach, the information need develops continuously, and the user goes through visceral need to compromised need in four phases of performing a task to solve a problem. The vertical interpretation, according to Cole, says "every user query has Taylor's four levels of information need," and the four information-need levels motivate both command- and question-based types of searches. Cole denies that information need is divided into stages and phases. Rather, it is the information search and information tasks that can be divided into phases. He believes that information need comprises levels, not phases, and that "Taylor has been misinterpreted to mean a phase approach or an evolution- or shifting information-need approach" (Cole 2011, p. 1229). Having analyzed Taylor's four-level information-need model, Cole (2011, p. 1229) proposes his hypothesis of information need, in which he believes "the information need manifests itself to the user in different ways," and as the user gets new information on the search process, the information need that is unconscious and visceral is brought into conceptualization. In his hypothesis, he "divides the performance of an information-based task (e.g., doing a project, solving a problem, completing an assignment) into Pre-focus, Focusing, and Post-focus phases or stages" and believes that the underlying information need does not instantiate fully to Taylor's deepest level of information need (visceral) until the user achieves focus (Cole 2011, p. 1229).

**Brookes** Brookes also can be viewed as one of the first proponents of the cognitive view. In his fundamental equation of information science, Brookes (1980), emphasizes the concept of knowledge structures, their interactions, and how the user's knowledge structure changes to a new one through the information he or she receives. This equation and its emphasis on knowledge structures and their interactions with each other, according to Belkin (1990) and Borlund (2010), demonstrate the power of cognitive perspective. Using a cognitive approach, Brookes (1980)

also explains the subjectivity and objectivity of information, i.e., when individual, subjective thought is expressed and recorded as accessible, it has been objectivized. However, as soon as this objective information reaches us, according to him, it becomes subjective.

**Dervin** Having a cognitive approach, Dervin's sense-making model is another attempt to reject the traditional information paradigm, in which information is viewed as an objective thing. Dervin's constructivist theory of information (1983a, 1992) views information not as something objective and external, but as something interpreted and constructed by the user through an internal, cognitive process of sense making at specific moments in time-space. As Dervin puts it, information is not seen as "a brick to be tossed" from an information system to the user, but rather as "clay to be molded and shaped" by the user who perceives the information (Dervin 1983a, p. 174). This suggests that information has a different meaning for each person in a specific context. This perspective challenges the notion that information is an objective entity that describes reality that users can receive. According to Savolainen (1993), "Sense-making theory, having its strongest roots in communication research, is a programmatic research effort suggesting user-centered ideas for the conceptualization of information seeking and use". In her Sense Making Model, Dervin (1983a, 1992) explains that information needs develop within a personal context in which the user can no longer go further through space or time without first gaining a new sense or better understanding of his or her situation. Such stops in the space-time continuum are caused by cognitive discontinuities, which Dervin refers to as information needing or gaps. Mackay (1960 cited in (Taylor 1967, p. 180)) calls this gap "a certain incompleteness in [user's] picture of the world" or in other words "an inadequacy in [user's] state of readiness to interact purposefully with the world around him". The gap, which the user seeks to cross, is the confusion and unanswered questions they have, whereas the ideas, thoughts, cognition and intuitions can be seen as the bridge that provides the means for continued movement in the user's context (Dervin 1983a). Moreover, Dervin considers sense-making as not only an intellectual and cognitive process, but also an emotional one. In sense-making, as a process of construction, the entire human (including body, mind, heart, and soul) and the whole experience of him/her, his/her feelings, including thoughts and actions, are involved and constantly evolve and intertwine with each other (Dervin 1998, Kuhlthau 1991).

**Belkin** Information in the user-centered approach is to be understood in relation to the user's knowledge structures in specific time-space when he or she recognizes some anomaly and incompleteness in his or her knowledge state. This recognized anomaly in the user's state of knowledge concerns some topic or situation, and generally, the user is unable to specify precisely what is needed to resolve that anomaly (Belkin et al. 1982). This is what Belkin (1980) terms the Anomalous State of Knowledge (ASK), which prevents individuals from going further in their rout and accomplishing their desired goals, resulting in a need for information to fix that anomaly in that person's knowledge structure.

**Ingwersen** Ingwersen (1982, p. 168) defines knowledge structures as a “system of categories and concepts”. According to him, “each individual’s image of the world consists of a conglomeration of different knowledge structures” with a variety of individual differences in these structures (Ingwersen 1982, p. 170). In his cognitive model of IR interaction, Ingwersen (1996) emphasizes that in IR, cognitive structures take the form of transformations that a variety of human actors or cognitive agents generate, including system designers and producers, IR-technique developers, indexing-rule constructors, indexers, authors, intermediary-mechanism designers, and users in a domain-related social or organizational context. Ingwersen and Järvelin (2005) view such divergence in cognitive structures as an asset and believe that what an effective IR needs is intense cognitive overlaps rather harmonization of those diverse cognitive structures. With his global model of polyrepresentation in IR, Ingwersen (1994) explains that polyrepresentation of user’s knowledge space means to represent not only the current information need, but also the underlying problem space, work task or interest, and the dominating domain. Ingwersen (1986) outlines three types of information need as verificative information need (VIN), in which the user wants to verify or locate items; conscious topical information need (CIN), in which the user wants to clarify, review, or pursue aspects of known subject matter; and muddled topical information need (MIN), in which the user wants to explore some new concepts or concept relations outside known subject matter. In another research, Ingwersen and Järvelin (2005) divide needs into two principal categories: specific and exploratory. They group searching for a known item, known data element, known topic or content, and factual data as specific needs, and muddled items, muddled data elements, muddled topics or content, and muddled factual data as exploratory needs.

**Wilson** Using a cognitive perspective, T.D. Wilson focuses on human information behavior. According to T. D. Wilson (1984, p. 197), the idea of knowledge structures, human perception, and cognition is the essence of the cognitive approach. Thus, a cognitive approach’s objective to seeking information, according to Wilson, is to discover how people’s knowledge structures, their images of the world, and their meanings of everyday life relate to available information, and how this image and knowledge structure determines their choice of information, then decide on what is relevant and how information can change these images and knowledge structures.

**Kuhlthau** Criticizing the traditional approach, Kuhlthau (1991) also emphasizes the necessity of investigating users’ cognitive thoughts and affective feelings, which are more difficult to observe, rather than study only physical actions taken during information seeking. Inspired by three theories – Kelly’s personal-construct theory, Taylor’s levels of information need, and Belkin’s Anomalous State of Knowledge - Kuhlthau (1991) suggests a series of six stages for the information-seeking process based on a series of several longitudinal user studies in which she examined not only the cognitive aspects, but also the feelings that users commonly experience. The six stages of her information-seeking process model (ISP) are initiation (recognizing information need), selection (identifying general topics), exploration

(investigating information on general topics), formulation (formulating a focus), collection (gathering information pertaining to focus), and presentation (completing information searches). Kuhlthau's ISP model incorporates both affective and cognitive aspects. During the series of stages, according to her research, feelings change during the construction phases, with thought changes in levels of information need, as well as changes in expression and mood. However, the ISP model has been criticized for being too general and not factoring in the searcher, work task, and search task as important dimensions in the information-seeking process (Pharo 2002). Moreover, it seems that her ISP model only can describe the information-seeking behavior of users who have just started to work on tasks that probably are new to them and outside their knowledge domains, in which users experience uncertainty in the first stage. For instance, identifying a "general topic" and investigating information on a "general topic" in the second and third stages of ISP are not applicable to all varieties of information needs and searches. Users do not always begin their searches by identifying "general topic"; sometimes they choose a very specific term or topic.

**Saracevic** In addition to cognitive and affective aspects, Saracevic considers situational aspects to be one of the most significant elements of the user side in an information-seeking process. In his stratified model of information seeking, Saracevic (1997) proposes three levels: cognitive (knowledge structure), affective (intent, beliefs, motivation, feelings, and desires), and situational (tasks) on the user side. Generally, the cognitive and user-centered paradigm focuses on what people really think, do, and feel when they search for information (Dervin 1998) in a particular situation. According to Borlund (2010, p. 23-24), the cognitive perspective acknowledges "the user's personal perception of the information need, the consequently subjective relevance assessments of information in response to that information need, and the context that surrounds the user, creates the given situation, and shapes the information need".

### **Relevance and user-centered approach**

Retrieval of information that is relevant to a user's problem or information needs is the principal objective of information systems and IR. Along with the two approaches of information retrieval, the concept of relevance can have two classes: 1) objective or system-based relevance that is logical and treats relevance as a static and objective concept, and 2) subjective or user-based relevance that considers relevance as a mental experience that involves cognitive restructuring (Swanson 1986, Borlund 2003). System-determined relevance has been criticized for being static and insensitive to the user and his or her cognitive states and situations. However, with user-centered relevance, the user is the ultimate judge, i.e., his or her evaluation is a cognitive phenomenon that depends on the user's perception of information needs, the situation, and time. Relevance is an implicit, if not explicit, notion in previous user-centered models of information behavior (e.g., Taylor's value-added model, 1986; Belkin's ASK model, 1982; and Dervin's sense-making model, (1983a)), in which the user needs to make judgments to determine whether the retrieved information will help resolve the information problem (Barry & Schamber 1998).

Having reviewed extant literature, Schamber et al. (1990, p. 774) drew three conclusions on the nature of relevance and its role in information behavior:

- Relevance is a multidimensional cognitive concept whose meaning is largely dependent on users' perceptions of information and their own information-need situations.
- Relevance is a dynamic concept that depends on users' judgments of the quality of the relationship between information and information need at a certain point in time.
- Relevance is a complex, but systematic and measurable, concept if approached conceptually and operationally from the user's perspective.

Regarding the concept of relevance, Patrick Wilson (1973, p. 458) differentiates among four kinds of relevance: psychological, logical, evidential, and situational. Schutz (1970) defines a system of relevances in which three basic and interdependent types of relevance – topical, interpretational, and motivational – interact dynamically. Within an IR context, Saracevic (1997) distinguishes among five manifestations of relevance: system or algorithmic relevance, topical or subject relevance, cognitive or pertinence relevance, situational or utility relevance, and motivational or affective relevance. As addressed by Barry and Schamber (1998), relevance can be characterized as being cognitive and subjective, situational, complex and multidimensional, dynamic, and systematic.

### 2.1.3 Motivation of information needs

Regarding the origin of information needs, Belkin (1980), takes a psychological approach. He believes that information needs form when a user recognizes an anomaly or inadequacy in their state of knowledge, often in respect to a particular problem that they do not know how to solve. Wersig (1971) in his article calls it “problematic situation” and Ingwersen (1992) renames it “the problem space of the user”. T. D. Wilson (1999) describes this problematic situation as follows:

Information seeking and retrieval are occasioned by uncertainty. [...] What is the cause of the uncertainty? A generalised answer is, 'a problem': the problem may be more or less recognisable as a problem in the normally understood sense of the word, but something in the individual's life-world, [...] has led [...] to a discrepancy between the typifications applied to the life-world and a phenomenon that, at first sight, cannot be fitted into those typifications. In other words, the individual is faced with a problematic situation (T. D. Wilson 1999, p. 265)

In the context of a problematic situation, based on Wersig and Wilson's analysis, a problem causes uncertainty and the state of uncertainty is the reason for requiring information. Accordingly,

Kuhlthau (1993a)'s Uncertainty Principle views information seeking as "a process of construction in which users progress from uncertainty to understanding". It is becoming aware of a lack of knowledge or understanding that causes the user to feel uncertainty and it is at this point when the user recognizes a need for information in order to reduce his or her uncertainty (Kuhlthau 1993a, Borlund & Dreier 2014). The concept of uncertainty can also be seen in Wilson's Problem Solving Model (1999) where the user moves from uncertainty to certainty as they progress through different stages of the problem resolution process, which includes problem identification (what kind of problem it is), problem definition (what the nature of the problem is), problem resolution (how to find the answer to my problem), and solution statement (the answer to the problem). The root of the 'problem' needs to be considered in order to understand what motivates individuals' information needs. White and Roth (2009, p.12) categorize motivations of information need as either internal or external and state that "the problem can be internally motivated (e.g. curiosity) or externally motivated (e.g. assignment)". An internally motivated information need can arise from curiosity, such as when an individual wants to learn more about something particular or to increase his or her knowledge in a specific area for personal development. An externally motivated information need can arise from something such as a work task that causes the individual to become aware of a gap in his or her knowledge structure.

T. D. Wilson (1999) believes that in every act of information seeking there are underlying task-related goals that need to be achieved. A task, as Toms (2011, p. 45) defines, has a "defined objective or goal with an intended and potentially unknown outcome and result, and may have known conditional or unconditional requirements" and states that the goals associated with the tasks are what trigger information needs and seeking. Toms (2011) analyses the task in the information seeking process more closely and explains that there are triggers and starting points for each task that may originate from another person or result from the demands of the job function. She also suggests an end or a stopping point for each task, which is when the goal associated with the task has been achieved or the task is abandoned because of different reasons such as the individual no longer needing the information, running out of time, encountering some information that leads "serendipitously to another task" (Toms 2011, p. 45). Byström and Hansen (2005) characterize tasks in three levels: information intensive work tasks, information seeking task, information search tasks (information retrieval task). To accomplish a single work task, they assert, a number of information seeking episodes may be involved and similarly, an information-seeking task may be composed of a number of search tasks. As Toms (2011) points out, the information search task is associated with searching an information retrieval system rather than other types of resource. The key activities of the information search tasks after recognizing the information need includes selecting a search system, formulating a query, executing the search, evaluating the results, and probably revising the query and reiterating the process until the information need is satisfied (e.g. Marchionini (1995)).

In general, all these tasks can have varying degrees of complexity. Task complexity, attributes and role in information needs and seeking has been a topic of research over the last three decades.

According to Byström (1999) and Campbell (1988), a complex task can be defined as one that requires a high cognitive and/or skill level from the task performer to be accomplished properly. However, Toms (2015) believes that there is no complex search task and referring back to Järvelin and Byström (Järvelin & Repo 1983, Byström & Järvelin 1995), he points out that the complexity of search task comes from the work task from which search tasks originates. Byström and Järvelin (1995) propose a categorization of tasks based on the 'a priori' determinability of tasks, which the extent to which the information requirements (what information is necessary to be searched), process (how to find the required information) and outcome (how to recognize the required information) of tasks can be described in advance. In the least complex tasks such as known-item search, the level of a priori determinability is high because the task performers are able to describe in detail and in advance what information is required, what processes need to be taken and what is the anticipated outcome. Whereas, in the most complex tasks like exploratory search tasks, these factors (inputs, process and outcomes) can not be determined a priori because of the dynamic nature of the user's perceived information need (Byström & Järvelin 1995, Byström 1999, Diriye et al. 2010). In short, the less sure a user is about task inputs, process or outcomes, the more complex the task is (Bell & Ruthven 2004).

Task complexity can be approached from different angles. Campbell (1988) classifies varying approaches as: (a) merely subjective and primarily a psychological experience; (b) an interaction between task and person; and (c) objective task characteristics. Subjective complexity is something that the task performer experiences. The complexity of a task, as Pharo (2004) states, is individually perceived. It is the subjective reaction of person to the task that happens based on person's familiarity to the task, person's short-term memory, person's motivation, person's uncertainty, person's attention, person's tenacity, time constraints and so forth (Campbell 1988, Pharo 2004). Person-task interaction complexity is what Byström (1999) calls 'perceived task complexity', which views the task complexity as being determined by the interaction between the task and task performer. According to this view, the characteristics of both the task and the person as the task performer are considered to determine the complexity of the task. However, task complexity can be also viewed objectively. In the objective task characteristics approach, task complexity is determined independently of the task performer. From the objective point of view, complexity of the task can be due to the existence of layers of task, which are progressively smaller sub-tasks that make up the task. Toms (2011) considers an inherently hierarchical structure for tasks, in that each sub-task has its own objectives, conditions, activities, actions and outcomes that are separate from the main task. In addition to hierarchical and serial relationships, Toms believes that there are also recursive relationships among sub-tasks, which result from the outcome from one sub-task leading to successive sub-tasks and the outcome of each sub-task is necessary for the main task to be developed and accomplished. Byström (1999) and Byström and Järvelin (1995) have suggested a list of the factors that determine complexity of tasks, which include: frequency or repetitiveness, analysability, certainty of action to be taken, variety, cognitive and skill demands, alternative paths of task performance, goals and conflicting dependencies, number of inputs, changes in tasks during performance, activity duration, and outcome novelty.

Considering the possible paths and the outcome, Campbell (1988) determines four attributes that contribute to task complexity: 1. multiple potential paths to the goal and to the desired outcome, 2. multiple desired outcomes, 3. conflicting interdependence among paths to multiple outcomes, and 4. uncertain or probabilistic links among path and outcomes. According to Campbell (1988, p. 43), each of these four characteristics “implies a high level of load, diversity, or rate of change” that increases the complexity of the task.

Given that, a PhD project can be considered a complex task with multiple potential paths to a goal that leads to several information needs.

### 2.1.4 Types of information needs

Types of users’ information needs can be approached from two perspectives: the required action and information-seeking process based on users’ cognitive status, and the nature and characteristics of the required information.

#### **From the perspective of the required action and information seeking process based on users’ cognitive status**

During the past few years, different researchers have tried to categorize users’ information needs by considering each user’s cognitive status and the extent to which the user is aware of his or her need and requisite information-seeking actions. For example, Ingwersen and Järvelin (2005) use the following three dimensions to classify information needs: 1) the intention behind the search task; 2) the type of current knowledge concerned with information seeking and retrieval and the underlying work task; and 3) the quality of what is known if it is well-defined or ill-defined. Marchionini (2006, p. 42) defines a hierarchy of information needs that ranges from “basic facts that guide short-term actions” to “networks of related concepts that help us understand phenomena or execute complex activities” to “complex networks of tacit and explicit knowledge that accretes as expertise over a lifetime”. However, all the defined categories and labels can be summarized between the two ends of an information-needs continuum, in which the unknown, uncertainty, and openness are at their highest levels on one end, while the lowest levels of the unknown, uncertainty, and openness lie on the other end, as shown in Figure 2.1.

The continuum’s two endpoints can be characterized by Kelly’s two attitudes (1970) that a user might assume regarding queries during the search process: “invitational” vs. “indicative”. In the invitational attitude, the answer that the user gets is not likely to be an exact answer to his or her question, but an answer to some other question that he or she has not yet thought to ask, and the user is “open for new ideas and interpretations, and receptive to change and adjustment according to what is encountered”. Conversely, the indicative mood causes the user to rely on the knowledge construct that he or she presently holds and to be resistant and reluctant to considering any new information and ideas (Kelly 2003; 1969). It is because the user knows,

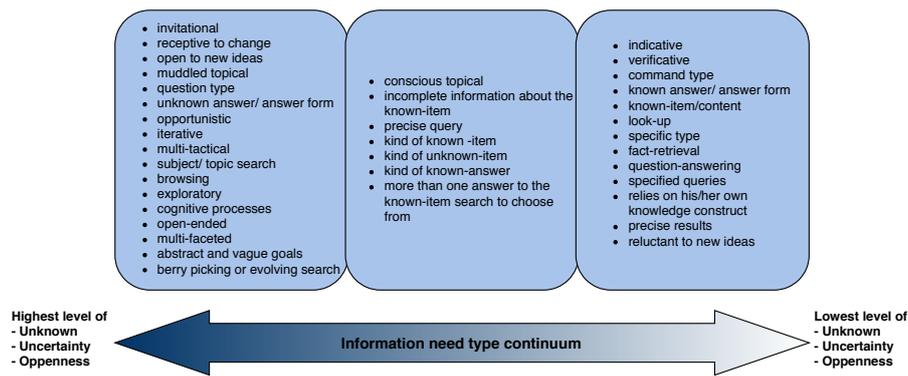


Figure 2.1: Information need type continuum

at least to an extent, about the item before searching for information on it and only wants that particular item and no other (Cole et al. 2010).

The endpoint with the lowest level of unknown, uncertainty, and openness is when the user knows what he or she is seeking. Under this endpoint, which Kelly (1970) calls “indicative” and Taylor (1968) labels “commands type”, the user not only knows what he or she is looking for, but also knows the answer, the form that the answer will take, or the information resource that holds the answer. It is when the user “knows exactly what he wants and can describe its form (book, paper, etc.) and its label (author and title)” (Taylor 1968, p. 191). Bates (1998) calls it a “known-item” search, as the user has specific knowledge of what he or she wants or has a particular item in mind when he or she starts to search for it (Wildemuth & O’Neill 1995). In other words, “the user knows exactly what is necessary to satisfy the need” (Belkin 1980, p. 137). Taylor describes it as a request for a specific item or specific subject combination that the user assumes will satisfy his or her need. This type of information need explains what Ingwersen and Järvelin (2005) call “specific type”, in which the user is looking for a known item, known data element, known topic or content, and factual data. It is comparable to a “verificative” information need that Ingwersen (1986) uses, in which the user wants to verify or locate items. Marchionini (2006) describes it as “look-up tasks” that need “analytical search strategies that begin with carefully specified queries and yield precise results with minimal need for result-set examination and item comparison”. Thus, it is a look-up about a known item or known answer form (Ingwersen & Järvelin 2005, Cole et al. 2010, Bates 1998).

Having the described continuum of information need in mind, Ingwersen’s “conscious topical” information need can be explained a little bit further in the continuum, as it is about clarifying, reviewing, or pursuing aspects of known subjects (Ingwersen 1986). The “networks of related concepts” in Marchionini’s hierarchies of information needs also can be placed here in the continuum. Besides, the “known item search 2” and “unknown item search 1” defined in Cole et al. (2010) can be viewed in the middle of the continuum. In his “known item search 2”, the user has a state of mind about the item, but his or her information about the known item is incomplete or inaccurate. While, in “unknown item search 1”, the user can describe what he or she is looking for and can formulate a precise query using the knowledge that a precise answer for his or her

question exists. However, the number of items that may include the answer is more than one in the results list, and this requires finding the best match.

The other endpoint of the information-need continuum comprises the highest level of unknown, uncertainty, and openness, entailing the user not knowing exactly what he or she should look for and where to start. The user might not even have any idea of what the form of the answer would be. When the user knows neither the answer nor its form, Cole (2011) calls it an “unknown item or unknown answer (or form of the answer) search”. As Belkin (1980, p. 137) describes, the user is “conscious of a need, but does not know what information would be appropriate to satisfy it”. According to Bates (1998), the user only knows fringes of a gap in his or her knowledge, making it extremely difficult to describe the information need and identify an effective start state for conducting a search in an information system (Cole 2011). This endpoint of the continuum, which Taylor (1968) labels “question type,” is when the user questions the information system with an ill-defined and muddled information need that is vaguely structured (Ingwersen 1992, Ingwersen & Järvelin 2005, Kim 2009, Wildemuth & Freund 2012) as a consequence of not having enough knowledge. This is the reason why the user wants to explore new things and concepts that Ingwersen labels a “muddled topical” information need (Ingwersen 1986). As the user explores new things and encounters new information, his or her information need is dynamic and gradually develops over time. Bates (1989) calls it “berry picking or evolving search”, in which the user starts with a broad, vague, and ill-defined information need, and as he or she gets new bits of information here and there, the user gets new ideas and directions to follow, and the query, information need, and desired outcome may change and evolve accordingly (Bates 1989; 2007). One of the principal characteristics of this kind of information need is uncertainty, in which the user has very abstract and vague goals (Aula & Russell 2009, Schacter et al. 1998). Therefore, users usually “submit a tentative query to navigate proximal to relevant documents in the collection, then explore the environment to better understand how to exploit it, selectively seeking and passively obtaining cues about their next steps” (White & Roth 2009, p.6). This explains why this kind of information need is characterized as open-ended, with many possible solutions targeting multiple information items (White & Roth 2009, Kim 2009, Wildemuth & Freund 2012, Schacter et al. 1998). Vakkari (2010) states that no identifiable correct answer exists for this kind of information need. Since it can be multifaceted and procedurally complex, including multiple sub-tasks, it may involve more than one search strategy and include iterative queries in different information-search systems and potentially multiple search sessions (Niu & Winter 2006, White & Roth 2009, Wildemuth & Freund 2012, He et al. 2008). As White and Roth (2009) note, it can last for days, weeks, or months. They also point out that this kind of information need is open-ended and multifaceted, following an information-seeking process that is “opportunistic, iterative, and multi-tactical” (White & Roth 2009, p.vi). This type of search also has been labeled a subject or topic search, browsing, and exploratory (Wildemuth & O’Neill 1995, Lee et al. 2006, Pirolli 2007) requiring cognitive processes, sense making, interpretation, comparing, and making qualitative judgments to help comprehend concepts and develop new knowledge (Marchionini 2006, Wildemuth & Freund 2012). White and Roth (2009, p.vi) also

admit that defining exploratory search as sense-making “aims to solve complex problems and develop enhanced mental capacities”. Some researchers consider learning to be a characteristic of an exploratory information need (Marchionini 2006, White & Roth 2009, Wildemuth & Freund 2012). However, learning can be a common characteristic in almost every kind of information need because we learn something every time we encounter a new piece of information, whether we need to do a quick look-up or a more exploratory search. However, the extent of how much we learn varies, but is greater with exploratory searches.

### **From the perspective of the required information’s nature and characteristics to satisfy an information need**

Types of information needs that users seek also have been studied from the perspective of the required information’s nature and characteristics. For example, Byström and Järvelin (1995, p.7-8), referring to other extant literature, classify types of information needed in tasks as “problem information”, “domain information,” and “problem-solving information”. They also characterize these information categories as follows:

- Problem information comprises “structure, properties, and requirements of the problem”. This information can be found in the problem environment or documents, in the case of old problems.
- Domain information includes “known facts, concepts, laws, and theories in the domain”, i.e., scientific information that can be found in published journal articles and textbooks.
- Problem-solving information entails “methods of problem treatment”, e.g., how the problem needs to be formulated, how to solve the problem, and what problem information and domain information should be used. They believe that such information typically is available only through experts and other knowledgeable people.

Hjørland (1997) describes a search typology in which he suggests the two principal non-document- and document-seeking categories. His non-document category, which also can be called fact retrieval, covers both facts and ideas (e.g., information about curing an illness) and information about physical things (e.g., information about locating a place). In the document-seeking category, he differentiates between known-item and unknown-item. Known-item entails verification of a known item, supplementary information about the item, or potential information about the item, whereas unknown-item entails identifying items with certain attributes, e.g., a specific language, or items related to a subject with the potential for either solving a specific problem or generally understanding a non-specific information need. Voigt (1961), in his report, discusses scientists’ approaches to information need in three different categories as follows:

- The current approach relates to keeping up to date with current progress in the field of scientists. It includes following current developments and recent research in relevant fields

of study. It could be more difficult when the scientist's subject field is a combination of different disciplines.

- The everyday approach is about finding specific pieces of information related to different parts of an experimental work or an understanding of a phenomenon in the work. Such small bits of information can be a piece of physical data, a specific method, a theory, a description of a substance or material, or an explanation of a specific observation. In this approach, according to Voigt, scientists want to go directly to the required information, and usually, the source of information is not that important, as long as the information is trustworthy. In fact, being easily available is of more concern than which source is the best.
- The exhaustive approach entails the scientist needing to find all existing relevant information on a specific topic. This approach usually is used at the beginning stages of a research project and when the scientist wants to prepare findings for publication and discuss a solution. In this approach, which Voigt calls a mechanical approach, the user tries to define the question as exactly as possible, and the machine provides all that exists in that database on the requested topic(s) and further selection is on the user.

In his final report on the process of answering reference questions, Jahoda (1977) identifies the different types of information that users sought as follows: dates, events, illustrations (pictures or symbols), numeric information (e.g., measurements, amounts), organizations, people, addresses or general locations, publications (e.g., citations, bibliographic data, etc.), and textual information (e.g., definitions, abstracts, recommended publications, general or background information).

Sridhar (2002), citing a Columbia University report (1960), lists the type of information that scientists mostly sought, including theoretical information, experimental results, data, methods, and procedural information. In a study on information-seeking behavior among scientists, Sami and Uplaonkar (2015) identified, through a quantitative survey, the types of information that scientists typically sought, which included current, research-and-development, factual, statistical, conceptual, retrospective, and socio-economic information.

Drawing on Grogan's (1991) schema<sup>1</sup> for types of reference questions asked in libraries, Duff and Johnson (2001) examined e-mail questions asked from archives and specified a new schema for questions types in archival domain. In their schema, they divided the questions into two main categories of limited-help questions<sup>2</sup> - including known item, fact-finding, material-finding, specific form- and open-ended questions including consultation, and user education. Hennieke (2017), inspired by Duff and Johnson's schema for type of questions asked in archives, studied

<sup>1</sup>He identified eight types of questions, falling into two broad categories: limited-help questions including administrative/directional, author/title, and fact-finding and open-ended questions including material-finding, mutable questions, research enquiries, residual enquiries, and unanswerable questions

<sup>2</sup>Their limited-help questions includes two more types of administrative/directional and service requests that are not in the scope of this section.

email reference questions of archives and classified users' questions into main categories of fact-finding questions and resource-discovery questions- that includes material finding, specific type, specific item, research questions, and consultation. (His third category of non-discovery questions that contains service request, administrative/directional, and user education questions is not relevant to the present study.)

## 2.2 Information organization

### 2.2.1 Cultural heritage information and increasing trend to be integratedly accessible

Different memory institutions not only hold different types of resources, but also organize and present information in different ways and with different levels of depth and details. Principal meta-data standards used for organizing cultural-heritage information include the Dublin Core Metadata Element Set (DC), Dublin Core Collections Application Profile (DCCAP), Machine Readable Cataloging (MARC), Metadata Encoding and Transmission Standard (METS), Metadata Object Description Schema (MODS), Categories for the Description of Works of Art (CDWA), VRA Core Record (VRA Core Record), Text Encoding Initiative (TEI), Encoded Archival Description (EAD), RSLP<sup>3</sup> Collection Description (RSLP), and CIDOC<sup>4</sup> Conceptual Reference Model (CIDOC CRM) (Lourdi & Papatheodorou 2006). According to Gill (2004, para. 4), although significant conceptual overlaps exist in the content and structure of information that archives, libraries, and museums curate, collections and descriptions of their content can vary according to factors such as “collection type, curatorial approach, subject discipline, granularity of description, level of detail of description, descriptive data structure, [and] descriptive data content values”. However, the information world has changed. Over the past few years, an increased focus on cooperation among archives, libraries, and museums has been experienced (Andresen 2006) to provide global access to their wealth of cultural-heritage content. Collaborations among memory institutions have created a global information infrastructure linking various resources from locations globally. Different information resources are distributed throughout archives, libraries, and museums that are related based on their underlying common history. According to Takahashi et al. (1998), the information is provided as a cross-cultural reference, e.g., one museum's information complements similar information at other museums or other memory institutions. Following the collaboration objective, an increasing effort has been made toward digitization of information among memory institutions. The rapid growth of digitized and digital cultural heritage information has elicited exponential growth in information

---

<sup>3</sup>Research Support Libraries Program

<sup>4</sup>Comité International pour la Documentation. CIDOC is one of the 30 International committees of ICOM (International Council of Museums). It is CIDOC rather than ICDOC because the committee was at its inception guided by mostly Francophones (French speakers) who named it the Comité International pour la Documentation. As the world has become more Anglophone (English speaking) the lingua franca of the committee has become English, and the de facto name of the committee has become the International Committee for Documentation. <http://network.icom.museum/cidoc/>

available online. On the other hand, users can be overwhelmed by the amount of information they receive after sending queries to information systems. It also is difficult for them to find relevant results based on their needs and preferences. Over the past decade, users' needs and expectations increasingly have become more sophisticated in the digital age. Memory institutions face many challenges as information resources become more technically complex, and those resources' users become more information-savvy. These memory institutions all share the goal of providing the most useful information possible to users in accordance with their needs, interests, and preferences. The invention of the web and semantic networks has elicited development of new models and ontologies in different areas. The cultural-heritage domain has been no exception and benefitted from these developments.

### 2.2.2 Semantic-network-based systems

Semantics, as pointed out by Carbonell (1970, p. 196), is the science of meaning: "A semantic information structure is an organization of units of information in terms of their meaning and mutual relationships". A semantic network is comparable to the general structure of human long-term memory and comprises a collection of nodes and arcs (links) interconnecting the nodes. Nodes represent objects, concepts, and events, and arcs (links) represent nodes' interrelations (Bic 1985, Quillan 1966). Semantic network systems have been developed in different fields with diverse goals. According to Barr et al. (1981), in psychology, Quillan (1966) developed the semantic network as an explicitly psychological model of human associative memory, while in the computer science and artificial-intelligence domains, semantic networks are a well-known approach to functional knowledge representations in information systems.

Quillan (1966, p. vii), in his thesis, develops a model for the general structure of human long-term memory in which "information about such things as the meanings of words is stored in a complex network, which then displays some of the desirable properties of a human's semantic memory". The central question of his research is: "What constitutes a reasonable view of how semantic information is organized within a person's memory" Quillan (1966, p. 1). His memory model comprises nodes that are interlinked by different kinds of associative links. The interconnected nodes in this network represent concepts. In his memory model, "ingredients used to build up a concept are represented by the nodes naming other concepts, while the configurational meaning of the concept is represented by the particular structure of interlinkages connecting those nodes to each other" (Quillan 1966, p. 14).

A couple of years later, using the semantic network as the basis of his system, called SCHOLAR, Carbonell (1970) provides an information network of facts, concepts, and procedures about geographical information to facilitate dialogue between students and computers with questions and answers from both sides. He believes that using a semantic network helps question-answering and two-way communication. The advantage of storing information in a semantic network, as Carbonell explains, is that the compatibility of organizational structures in the human mind with information systems facilitates communication between man and machine and provides

more effective information retrieval accordingly. According to Carbonell (1970, p. 197), “the organization in the computer provides retrievability according to the dimensions that the users consider relevant”. He assumes that human symbolic memory is organized like a semantic network, and if an information system uses such a semantic network to organize information, the information system will work with the same kind of information structures that the human mind uses to function.

### 2.2.3 Ontology vs. semantic network

Semantic networks and ontologies might look similar, but they are not the same. Semantic networks (Quillan 1966) were developed earlier than ontologies, and they lack formal semantics and precise semantic characterization despite the fact that they have “semantic” in their name. Ontologies have well-defined and formal semantics. Ontologies’ rich, logical structure cannot be compared to a semantic network. Sowa (1991, p. 1) describes a semantic network as a graphic notation used to represent knowledge “in patterns of interconnected nodes and arcs” that can represent hierarchical (superclass/subclass) relationships and instances for each class, but not the restrictions on these relationships and property characteristics.

Ontology, as “an explicit specification of a conceptualization” (Gruber 1993, p. 1), is an explicit, formal description of common concepts in a domain, with each concept’s properties describing various features and attributes of the concept, as well as restrictions on properties (Gruber 1993, Noy et al. 2001). Through enhancements in data sources’ explicit semantic information (Hajmoosaei & Skoric 2016), ontologies provide more enriched meaning. An ontology, as Agarwal (2005) defines it, is the manifestation of a shared understanding of a domain that facilitates effective communications of meaning, consequently boosting inter-operability of information systems. Since entities and properties in ontologies characterize the real world and explicitly represent domain semantics, they can span heterogeneously structured databases with comparable semantics (Obrst 2003) and make domains’ knowledge reusable.

In the cultural-heritage domain, as Gruber (1993, p.11) explains, a bibliography ontology’s purpose is to support knowledge-sharing tasks, which Gruber says entail “exchanging bibliographic data among databases, integrating bibliographic databases with other databases (e.g., address books, company directories), linking citations across hyperdocument boundaries, and providing network-based services for bibliographic data processing”.

### 2.2.4 Organizing strategies and ontologies

Meeting users’ changing needs and expectations has necessitated the use of complex technologies to organize and represent knowledge across memory institutions. According to Hjørland (2013), “Knowledge organization (KO) is about classifying knowledge, for example, to define concepts and determine their semantic relations”. Hjørland continues: “KO is about concepts and their semantic relations (and at the same time about the real world)”. As Sowa (2000, p. xii) and

Oldman (2014) note, knowledge representation is the “application of logic and ontology” to represent real-world entities in a form that computers can understand.

In organizing knowledge and providing integrated access to cultural heritage in formations distributed across archives, libraries, and museums, i.e., memory institutions, it is important to consider the challenge of data heterogeneity and compatibility among these institutions’ existing information systems. Different memory institutions not only hold different types of resources, but also organize and present the data in different ways, with different depth and detail levels. Considering memory institutions’ diversity, ontologies as a solution support semantic interoperability and integration. Ontology-based metadata provides more efficient retrieval and access by supporting, for example, “very precise queries,” “explicit paths from vague to precise queries,” “integrating catalogs,” etc. (Weinstein 1998, p. 255). As Coyle (2015) notes, flexibility and extensibility are a bibliographic standard’s key aspects. Among different efforts, the CIDOC Conceptual Reference Model (CRM) is a rich ontology that provides “a real-world, empirically based representation” (Oldman 2014, p.1) and offers an alternative solution to the challenge of providing meaningful integrated access to heterogeneous cultural-heritage information across archives, libraries, and museums.

### **Entity-relationship models**

Entity-relationship models have been proposed for use in information modeling and analysis. Most are based on the concepts of entity, relationship, and attribute. In 1976, Chen (1976) proposed the entity-relationship model in his article, titled “The entity-relationship model: Toward a unified view of data”. He believes this model “incorporates some of the important semantic information in the real world” and can be used “as a basis for a unified view of data” (Chen 1976, p. 10). In entity-relationship models, an entity is a thing that can be identified distinctly. The entity defines the model’s framework and functions as a node, while the relationship is the association among entities and shows how the entities connect and interact with each other. Attribute depends on entity and provides information about the entity and characterizes specific instances of the entity (Riva et al. 2017, Chen 1976). Functional Requirements for Bibliographic Records (FRBR) is an example of an entity-relationship model that identifies and clearly defines the entities of interest to users of bibliographic records, the attributes of each entity, and the types of relationships that operate among entities (IFLA study group on the Functional Requirements for Bibliographic Records 1998).

**FRBR and its family of models** Functional Requirements for Bibliographic Records, as an entity-relationship model, was developed by a study group that the International Federation of Library Associations and Institutions (IFLA) appointed during the 1991-1997 period and was published in 1998 to represent a “generalized view of the bibliographic universe” (IFLA study group on the Functional Requirements for Bibliographic Records 1998). FRBR “is oriented to achieve a more holistic approach to retrieval and access from a user’s perspective” (Giannoulakis

et al. 2018, p. 640). According to Doerr (2009, p. 471), FRBR's innovation is "to cluster publications and other items around the notion of a common conceptual origin – the Work – in order to support information retrieval and to initiate a new bibliographic practice". FRBR's key concepts are represented through 10 entities that have been divided into three groups, the first of which represents intellectual or artistic products that are described primarily in bibliographic records, comprising Work (intellectual or artistic creation of the mind, a set of concepts), Expression (intellectual or artistic realization of a Work, a set of signs), Manifestation (one of the possible embodiments of an Expression), and Item (a single exemplar of a Manifestation); The second group of entities represents agents who are responsible for the intellectual or artistic content and its production, or even custodianship, comprising Person and Corporate Body. The third group represents subjects of intellectual or artistic products (i.e., the entities that have a "subject relationship" to the Work entity), comprising all the entities already declared in Group 1 and Group 2, plus Concept, Object, Event, and Place (IFLA study group on the Functional Requirements for Bibliographic Records 1998, Žumer & Le Bœuf 2006). However, the model operates at a conceptual level and maybe not at the level required for a fully developed data model (IFLA study group on the Functional Requirements for Bibliographic Records 1998).

Over a decade after FRBR was developed, IFLA has led the development of two extensions to FRBR that include authority data because FRBR focused particularly on bibliographic data. Functional Requirements for Authority Data (FRAD), as a conceptual model for authority data, was developed in 2009 to focus on agents and works, while Functional Requirements for Subject Authority Data (FRSAD), as a conceptual model for subject-authority data, developed the subject relationship and was approved in 2010. These three models are known as the FRBR family of conceptual models.

**IFLA Library Reference Model (LRM)** To resolve inconsistencies among the three FRBR family models, the IFLA Library Reference Model (IFLA LRM) was developed in 2017 (Riva et al. 2017) within an entity-relationship modelling framework that unifies the separately developed conceptual models of the FRBR Family: FRBR, FRAD, and FRSAD. The IFLA LRM model is a single, complete, and logically consistent model of the library domain that covers all aspects of bibliographic information and removes "barriers to adoption due to divergent conceptualizations between the models of the FRBR Family" (Žumer & Riva 2017, p. 13). LRM aims to be a general, high-level conceptual reference model designed to be expanded for implementation.

### **Object-oriented models**

Relational databases are fast and reliable for massive quantities of data, but they are best-suited for large quantities of highly standardized data and likely not for representing the complexity of bibliographic information (Weinstein 1998) in the cultural-heritage domain. Bibliographic data, as Weinstein (1998, p. 255) explains, "requires many descriptive dimensions, partial description at multiple levels of granularity (with any combination of dimensions), and viewing from many

perspectives (access by different sequences of dimensional values)". Object-oriented databases are an appropriate solution for complex domains and the challenge of representing various descriptive dimensions and granularity levels. Since the 1990s, both in computer science and in the cultural-heritage domain, a shift in approach has occurred from entity-relationship models to object-oriented models. Nunes (2001, p. 30) describes object orientation as "a technique for organizing knowledge, that is, to organize the way we think about the world". He says this organization of knowledge "is expressed in terms of types of things (object types), their attributes, operations, relationships, rules that drive their behavior and collaborations". The turn from entity-relationship models to object-oriented models in the cultural-heritage domain began in the mid-1990s with the development of CIDOC CRM in the museum community. Some of the motivating factors to shift toward object-oriented modelling techniques, as Doerr and Crofts (1999) and Oldman (2014) point out, are as follows: the object-oriented models support a semantically richer form of representation than entity-relationship models; the more meaningful form of representation and the mechanism of specialization make the object-oriented models more readily and sustainably extensible than entity-relationship models; the specialization and aggregation of entities provides a means to present variable levels of granularity without showing the complexity (because it allows for removal of redundant representation); and the object-oriented models provide the ability to represent a range of generalization and specialization. These all make the model more flexible and adaptable, which is useful for developing interactive representation and visualization in integrated systems.

The object-oriented model offers uniform, extensible, and reusable data, and reduces the semantic gap between the information system and real-world concepts (Vlahavas & Bassiliades 1998). According to Doerr and Crofts (1999), the object-oriented reference model provides a means to define "the semantic values of data structures with the precision needed to ensure reliable communication and mediation of cultural information". Hence, CIDOC CRM and its extensions, like the FRBRoo model, could be used as the basis for implementing cultural information systems. The next section presents some basic information about these models.

**CIDOC Conceptual Reference Model (CRM)** While the library community was developing the FRBR entity-relationship model in the 1990s, a similar phenomenon occurred in the museum community. The ICOM CIDOC developed an object-oriented conceptual model called CIDOC Conceptual Reference Model (CRM) for the information that museums produced to document their collections. The first complete edition of CIDOC CRM was published in 1999, then in 2006, it was issued as ISO standard 21127<sup>5</sup>. Although the CIDOC CRM developed within the museum community, it has been designed with a broader goal of facilitating rich information exchange among archives, libraries, and museums (Gill 2004, Le Boeuf 2003, Doerr et al. 2007)

CIDOC CRM is a complex thing. Its present version (Version 6.2.3, 2018) comprises 94 classes

<sup>5</sup><https://www.iso.org/standard/57832.html>

representing general notions in the domain, such as CRM class, E21 Person, and 174 Unique properties, which represent binary relations that link classes, such as the CRM property, P152 has parent. Classes and properties are used to express ontological knowledge via various kinds of constraints, such as subclass/subproperty links (e.g., E21 Person is a subclass of E20 Biological Object) or domain/range constraints (e.g., the domain of P152 has parent is class E21 Person) (Le Boeuf et al. 2018). Since the CIDOC CRM is event-based, Temporal Entity (E2) is a central notion at this model's core. The Temporal Entity happens within a Time Span (E52) and at a Place (E53), and it "brings about some change in the world" (Žumer & Le Boeuf 2006). Some Actors (E39) participated in this Temporal Entity, which affects or refers to Conceptual Objects (E28) or a Physical Thing (E18). The CIDOC CRM lets us name anything and apply names to things, using Appellation (E41). Finally, to categorize all these things, CRM uses the notion of Types (E55). These entity types form the upper level of the CIDOC CRM, and everything else in this model is a specialization of this top level (Oldman 2014). In CIDOC CRM, entity types and relationships exist in a hierarchy of meanings. Both classes and properties can have specialization (sub-class/sub-property) and generalizations (super-property/super-property). Since the structure of CIDOC CRM is poly-hierarchical and not flat linear, it provides different levels of generalization or specialization in the optimal range. This characteristic is important because we cannot always be precise about everything that we want to describe, but when we have more specific information, we should specify more (Oldman 2014).

Employing object-oriented, data-modeling techniques, the CIDOC CRM provides a formal expression of the basic concepts and their relationships behind the structure of the various data that are supposed to communicate (Doerr & Crofts 1999, p.1). The CIDOC CRM provides "a model of the intellectual structure of cultural documentation in logical terms" by formally describing the implicit and explicit relationships of cultural-heritage concepts. It lets us explain how apparently different entities are intellectually interconnected (Le Boeuf et al. 2018, p. i). By promoting a common understanding of cultural heritage information, the CIDOC CRM is intended effectively to structure and relate, information, assets of cultural documentation, support associative queries against integrated resources, and provide a basis for the semantic integration and interoperability between heterogeneous data sources regardless of subject matter and their applied classifications (Le Boeuf et al. 2018, Oldman 2014). To facilitate semantic integration and interoperability, significant effort has been invested in mapping different models and standards used in culture heritage institutions to CIDOC CRM as the core ontology (e.g. Le Boeuf, 2006; Kakali, 2007; Stasinopoulou, 2007) and also mapping information into the CIDOC CRM through interpreting and encoding the semantic content of cultural heritage documentation using CIDOC CRM ontology as a formal ontology (e.g. Holmen, et al. , 2004).

CIDOC CRM, as a general model, currently has included several domain-specific or application-specific extensions, such as FRBRoo<sup>6</sup> (Functional Requirements for Bibliographic Records),

---

<sup>6</sup><http://www.cidoc-crm.org/frbroo/>

PRESSoo<sup>7</sup> (modeling of bibliographic information pertaining to continuing resources), CRMinf<sup>8</sup> (argumentation model), CRMarchaeo<sup>9</sup> (excavation model), CRMsci<sup>10</sup> (scientific observation model), CRMgeo<sup>11</sup> (spatiotemporal model), CRMdig<sup>12</sup> (model for provenance meta-data), CRMba<sup>13</sup> (model for archaeological buildings), and CRMtex<sup>14</sup> (model for the study of ancient text).

**FRBR object-oriented (FRBRoo)** When the two models, FRBR in the library community and CIDOC CRM in the museum community, were completed, they started talking about combining their tools and resources. They recognized the benefits of harmonizing the two models and to provide a common view of cultural-heritage information to support information integration and interoperability. The harmonization idea first was expressed in 2000 during the European Library Automation Group's (ELAG) 24th Library Systems Seminar in Paris. The idea gained steam in subsequent years, leading to formation of the International Working Group on FRBR/CIDOC CRM Harmonization in 2003 by both communities' representatives. This effort's outcome was published in 2009 as the first version of FRBR object-oriented (FRBRoo, v. 1.0), which was an object-oriented version of the original FRBR model. The second version (FRBRoo, v.2.4) was published in 2015 and is the object-oriented interpretation of the FRBR family of conceptual models, including FRBR, FRAD, and FRSAD. It is not the same as any of these three existing entity-relationship models. It does not change the concepts expressed in the original three models, nor does it transform them into something conceptually different, but rather expresses the conceptualization of the FRBR family using the object-oriented methodology, instead of the entity-relationship methodology (Bekiari et al. 2015). In other words, it provides an alternative perspective of the FRBR family of models. The objectives of FRBR are reaching a common perspective on cultural-heritage information, enabling data-integration interoperability between the library and museum domains through a shared ontology, and extending the scope of both the FRBR family of models and CIDOC CRM beyond their original domain, giving the library community the opportunity to benefit from further extensions of CIDOC CRM. Beyond that, extending the scope of the CIDOC CRM to bibliographic records in formation paves the way for more extensions to other domains and formats, such as EAD, TEI, MPEG7, etc. (Bekiari et al. 2015).

FRBRoo makes some concepts and relationships explicit that were implicit in the original models. Bibliographic records contain some implicit information about their complicated path from when the idea forms in its creator's mind to the physical item in a user's hands (Coyle 2015). Unlike FRBR family models that don't cover this information, the FRBRoo model makes this implicit

<sup>7</sup><http://www.cidoc-crm.org/pressoo/>

<sup>8</sup><http://www.cidoc-crm.org/crminf/>

<sup>9</sup><http://www.cidoc-crm.org/crmarchaeo/>

<sup>10</sup><http://www.cidoc-crm.org/crmsci/>

<sup>11</sup><http://www.cidoc-crm.org/crmgeo/>

<sup>12</sup><http://www.cidoc-crm.org/crmdig/>

<sup>13</sup><http://www.cidoc-crm.org/crmba/>

<sup>14</sup><http://www.cidoc-crm.org/crmtex/>

information explicit and consequently facilitates moving “from conceptual model to real-life applications” (Bekiari et al. 2015).

**CRM inf: Argumentation model** The argumentation model CRMinf uses and extends the CIDOC CRM (ISO21127) as a general ontology to support integrating meta-data about argumentation and inference making in descriptive and empirical sciences. It aims to facilitate “the management, integration, mediation, interchange, and access to data about reasoning by a description of the semantic relationships between the premises, conclusions, and activities of reasoning” (Stead et al. 2015, p. 5).

### 2.2.5 Ontologies’ compatibility with users’ mental models

To implement ontologies in information systems efficiently, we first must conceptualize users’ mental models and find out how users search for information and perceive entities and relationships between entities in the bibliographic universe. Mental models are constructed in the mind, the result of intuitive perception, knowledge, imagination, and comprehension. Mental models are formed and evolved through individuals’ interaction with the surrounding world. As Norman (2014, p. 7) explains:

people’s views of the world, of themselves, of their own capabilities, and of the tasks that they are asked to perform, or topics they are asked to learn, depend heavily on the conceptualizations that they bring to the task. In interacting with the environment, with others, and with the artifacts of technology, people form internal, mental models of themselves and of the things with which they are interacting.

Specialists usually spend a considerable amount of time and effort developing complex ontologies. However, ontologies have been criticized for not being based on explicit user studies. We do not know that much about users’ mental models of the bibliographic universe yet. As Žumer et al. (2012, p. 149) note, “we still have little research-based evidence on exactly what attributes and relationships different user groups require”.

To see whether ontologies are in line with users’ mental models, Žumer and her colleagues have conducted user studies (Pisanski & Žumer 2010a;b, Žumer et al. 2012, Budanovic & Žumer 2015, Noc & Žumer 2014) with different groups of users and found that “FRBR is intuitive for users” (Budanovic & Žumer 2015). In one of the studies (Pisanski & Žumer 2010a;b), they tried to see whether FRBR’s structure would be confirmed by non-librarians’ mental models through card sorting, concept mapping, and comparison tasks. Their study justified the FRBR distinction between manifestation and expression. In another study, Budanovic & Žumer (2015) performed a similar study using concept mapping to see whether cataloguers’ mental models resemble the FRBR conceptual model. Their findings show that most cataloguers’ mental models were close to FRBR, although none was exactly like FRBR. In a pilot study

using card sorting and concept mapping as methods, Noc and Žumer (2014) compared the bibliographic universe's mental models with music resources for the FRBR model. The pilot study's results show similarity between users' mental models and FRBR's structure of work, expression, manifestation, and items. Focusing on derivative relationships in the bibliographic universe, Tallerås et al. (2018) studied non-experts' (novice students) conceptualizations of the "relationships between independent, but strongly related, entities" using a concept-mapping method. Their findings show that despite the variety in participants' conceptualizations, they grouped them into two categories: some who use what Tallerås et al. call "single-entity models" and many who use "multi-entity models," in which they assume high-level nodes to collocate related documents. Considering users' own information needs, Hennicke (2017) tried to find a modeling of archival user needs. He investigated common patterns in archival user enquiries and represented them in the CIDOC CRM model to improve the discovery of facts and materials in archives. He also proposed some new classes and properties to the CIDOC CRM, such as documentation activity, self-documentation, documentation of others, and mandates.

The present study's focus is not on data presentation explicitly, but we hope that the study's findings can shed light on users' actual information needs and can be used to design ontology-based information systems' presentation layers. Regarding the design of data presentation in ontology-based systems, some effort has been devoted to users' preferences. For example, Aalberg et al. (2016) developed a search application called BIBSURF to experiment with search, ranking, and filtering of bibliographic data that are organized according to the Library Reference Model (LRM). They use this system to see how specific bibliographic patterns need to be presented according to users' preferences, particularly what types of entities users prefer to be presented in search-results lists. Having built on the concepts of work family in FRBR, Merčun (2012) proposes a model to support a user-friendly presentation of bibliographic data in FRBR-based information systems. In her proposed model, she tried to bring together and present all versions of a work, related works, and other works by and about the author. She also conducted two user studies to evaluate the model and different information-visualization designs.

**Conclusion** Having this assumption that ontologies provide efficiency in describing information and relationships among it, the question should be how much of this interrelated cultural information should be displayed to the user and how much should be retrieved for each query? In other words, how much of this semantic information would the user need to have represented to meet the user's information needs? We need more empirical user studies with an ontological approach to bring cognitive structures of system designers and ontology developers into accord with those of users. Although extant user studies have shed light on different aspects of this complex concept of information needs, they have failed to represent actual information needs adequately as wholes and users' perspectives on interconnected cultural heritage information in a natural setting. Thus, this exploratory study tries to conceptualize users' information needs in a setting that is as natural as possible to see how users' information needs can be projected into ontologies and to identify common patterns in their needs.



## Chapter 3

# **Methodology**

*A qualitative researcher should be curious, creative, and not afraid to trust his or her instincts.* (Corbin & Strauss 2008, p.16)

This chapter discusses the methodology (naturalistic inquiry) and the method (contextual inquiry) selected for this exploratory study. It starts with a brief discussion of the naturalistic approach and a description of contextual inquiry as the research method, then continues with a description of the setting and the steps taken to advance the study, including sampling (its type and size, the processes used for recruiting the selected sample, and the demographic details of the recruited participants), data collection, and data-analysis stages. The chapter ends with a discussion the trustworthy and credibility of the research and its findings and limitations of the study.

### 3.1 Naturalistic Inquiry

The present study's principal concern is to investigate users' information needs when they search for cultural-heritage information. The researcher wanted data that more closely reflected users' real, lived experiences. Since naturalistic inquiry approximates natural and uncontrived conditions (Wildemuth 2016), it was the best approach through which the researcher could address my research questions. With a positivistic approach, reality is viewed as a single and observable entity that can be divided into specific variables to be studied, manipulated, controlled, and predicted. However, with a naturalistic approach, reality comprises multiple perspectives that individuals subjectively construct, and the researcher does not control, manipulate, or impose constraints on the outcome of what is being studied (Wildemuth 2016, Jacobs 1985, Merriam & Tisdell 2015). With a positivistic approach, which is scientifically objective and rational, the researcher and participant or object of inquiry are independent, and the researcher can develop context-free generalizations. However, with a naturalistic approach, the researcher and participant or object of inquiry are mutually dependent, and the focus of inquiry is on meaning in context (Wildemuth 2016, Merriam & Tisdell 2015). Unlike with quantitative methods, in which researchers deductively test a hypothesis or theory, the process in qualitative research is inductive, and the researcher gathers data to build concepts, hypotheses, or theories. For decades, scientists using positivistic (rationalistic) approaches conducted laboratory and experimental studies to investigate users' information behavior in a way that can be quantified and statistically analyzed. However, laboratory studies in particular, as Wildemuth (2016) explains, fail to:

- Gather a detailed, unprejudiced record of people's behaviors, beliefs, and preferences
- Explore people's behavior in the context of their own work and life
- Intensively observe particular contextual elements, such as settings and artifacts
- Uncover tacit meanings and understandings common in communication and social interaction

To address this gap, researchers developed and used a variety of naturalistic research methods, using different degrees of naturalism, to study users in their natural settings, either by going into the field to observe users in their individual environments or by replicating elements from these natural environments elsewhere (Wildemuth 2016). Think-aloud protocols have been used to understand users' cognitive processes based on their verbal reports on their thoughts while interacting with systems (Ericsson & Simon 1993). However, the difficulty that people have in reporting their cognitive processes presents a limitation (Nisbett & Wilson 1977). The transaction-log analysis method is not limited, i.e., dependent on the participant's memory and his or her ability to describe the interaction, but it lacks context, so we do not know about the participant's intentions, expectations, or satisfaction levels (Wildemuth 2016). Reference-question analysis may benefit from information that the participant provides, e.g., intentions, known information, and information needs. However, it fails to study information need as a whole. Reference-question analysis focuses on formalized needs, not on other information-need levels. The method also lacks the benefit of the user's presence for further explanations and clarifications. The naturalistic paradigm, which Lincoln and Guba (1985) address, suggests that realities are wholes that cannot be understood apart from their contexts, nor can they be fragmented for separate study of their parts with simultaneously emphasis on the whole. They contend that the whole is more than the sum of its parts. Users' information needs are a subjective, cognitive reality that individuals construct while interacting with their environments. The only way to understand this reality deeply is through qualitative research, in which the researcher gets involved in participants' reality and interacts with them meaningfully (Bloomberg & Volpe 2018). According to Corbin and Strauss (2008), the most important reason for choosing to do qualitative research is the desire to step beyond the known and enter study participants' realities to view the world from their perspective, and in doing so, make discoveries that will contribute to the development of empirical knowledge. Thus, contextual inquiry is the best method for investigating users' actual information needs in their natural settings.

## **3.2 Study's design**

Qualitative research, according to Saldaña (2014), is a customized, inductive, and emergent process in nature that allows the researcher to graft his or her personal signature on the study's design, implementation, and write-up. Research methodology, as Buckley et al. (1976) define it, is a "strategy or architectural design by which the researcher maps out an approach" to seek answers to the problem at hand. Following the naturalistic approach, the present study aims to explore users' (PhD candidates) real information needs through contextual inquiries.

### **3.2.1 Contextual Inquiry**

Contextual inquiry is a qualitative field-research method and, according to Holtzblatt and Beyer (2014), the first phase of contextual-design methodology to design complex systems optimally

based on users' requirements. Contextual inquiry essentially comprises going out into the field where users work, observing them working, and talking to users about their work and experiences (Holtzblatt & Beyer 2014, Beyer & Holtzblatt 1999, Raven & Flanders 1996). In other words, contextual inquiry is based on observations of, and conversations with, users in the context of their work. Watching users engaged in real activities helps researchers uncover details about users' activities and reveals aspects that users may not be able to articulate (Holtzblatt & Beyer 2014). To conduct contextual inquiry, according to Beyer and Holtzblatt (1997), four principles need to be considered<sup>1</sup>:

- **Context:** Go to users' workplaces, observe them working as they usually would, and discuss what they are doing and why. Observing and talking to participants in the context of performing routine, specific tasks help the researcher gather rich data that differ from the type of data obtained from questionnaires or other more quantitative methods.
- **Partnership:** Collaborate with users to uncover unarticulated aspects of their work, eliciting information in a natural and relaxed way.
- **Interpretation:** Develop a shared understanding with users about the meaning of their words, emotions, and actions. Rather than interpret such data on our own, we can ask participants why they did things and what they meant by these actions (Raven & Flanders 1996). Such a partnership helps the researcher and informant explore issues together.
- **Focus:** Direct the inquiry by concentrating on relevant aspects while observing.

The contextual-inquiry process includes semi-structured interviews and observation sessions. A semi-structured interview method is used to obtain information about the context of use, in which users first are asked a set of standard questions, then observed and questioned while they work in their own environments. Because users are interviewed in their individual environments, the analysis data are more realistic than data obtained in a laboratory environment. Both interviews and observations have advantages and disadvantages when used alone, but when they are used together in contextual inquiry, they can provide multidimensional perspectives on users' needs. Having received inspiration from contextual inquiry, the researcher collected data through semi-structured interviews accompanied by observations in a natural setting as much as possible. The researcher learned that the best way to observe how users search for cultural-heritage information, and determine what users really need, is to talk to them and ask questions while observing their search behavior in natural settings. As Kvale and Brinkmann (2009, p. xvii) also point out, "If you want to know how people understand their world and their lives, why not talk with them?" In-depth interviews entail purposeful conversations that enable the researcher to ask questions directly based on the research study's goals and obtain rich data about people's thoughts and rationale that underlie their behavior (Saldaña 2014, Wang 1999). Semi-structured

---

<sup>1</sup>Raven & Flanders (1996) considers only three principles for contextual inquiry that include: context, partnership and focus

interviews, also states, reveal a single individual's worldview, providing freedom to explore views and opinions in a more detailed way. While the semi-structured format ensures that specific key questions and topics are addressed in every interview, it also allows the researcher to ask additional questions when appropriate. Observation also is a systematic data-collection approach in which the researcher observes people engaging in normal activities. It helps the researcher observe what people actually do, rather than what they say they do, using his or her senses to examine people in natural settings or in naturally occurring situations. Observation is an essential part of gaining an in-depth, rich understanding of participants' behavior in naturalistic settings, in which the researcher is granted access to the context and meaning surrounding these phenomena. Using observations alongside interviews, as part of contextual inquiry, helps both the participant and researcher examine themselves in natural situations and process their thoughts more openly, with less pressure to recall what occurs in everyday situations. To capture such interview data more effectively, recording these interviews is considered an appropriate practice, making it easier for the researcher to focus on an interview's content and verbal prompts, as well as enabling the transcriptionist to generate a verbatim transcript of the interview (Jamshed 2014). One of this user study's most significant strengths, methodologically speaking, is that users are observed in their natural settings performing their own natural tasks and demonstrating feasibility and in-context usefulness. By providing rich data, the selected method helps take a deeper look into users' information needs in a rather natural setting.

### 3.2.2 Sampling

In contextual inquiry, you must select your informants carefully to ensure that they represent an important audience. Since this study's objective is to explore users' information needs from different fields, study participants were selected purposefully. Purposive sampling is different from random or representative sampling usually used in quantitative studies. Purposive sampling intends to maximize the range of information collected. To ensure that participants regularly are engaged in search tasks and have cultural-heritage information needs, emphasis was placed on recruiting PhD candidates because conducting PhD research presumably creates information needs, eliciting searches for information from cultural-heritage resources. PhD candidates are assumed to be researchers with complex tasks, who are actively searching for information. To ensure that this study's research questions were answered fully, it also was necessary to use informants who would search for information within different cultural-heritage information institutions, including libraries, archives, and museums. Operating under the assumption that almost all PhD candidates would need to use library materials to work on their projects, the focus was on selecting informants who were at least potential users of museum objects and/or archival materials. Selection criteria also included seeking participants from a variety of research domains and fields of study, striving for a balanced gender mix, and aiming for variation in participants' PhD project stages.

Sampling criteria for this study included:

<i>Participants</i>			
<b>Gender</b>	<b>Age</b>	<b>Department/Institute</b>	<b>Research stage</b>
Female	50-60	Department of Archivistics, Library and Information Science	ending
Female	30-40	Department of Archaeology, Conservation and History	beginning
Female	30-40	Department of Sociology and Human Geography	middle
Male	40-50	Museum of Cultural History	middle
Female	30-40	Department of Literature, Area Studies and European Languages	middle
Female	30-40	Department of Literature, Area Studies and European Languages	middle
Female	50-60	Museum of Cultural History	middle
Male	30-40	Department of Public and International Law	beginning
Female	30-40	Department of Archaeology, Conservation and History	ending
Male	30-40	Museum of Natural History	beginning

Table 3.1: Profile of the participants

- Active searcher (emphasis on PhD candidates)
- Potential users of ALM (archive, library, and museum) resources
- Variation in disciplines

To recruit participants, several web resources were used, including the university's website, museums' websites, different faculty webpages, and PhD candidates' profile pages, which include information on their PhD projects, e.g., titles or topics, and sometimes including short descriptions of their projects, research interests, published work, and recent page updates. PhD candidates who did not have informative web pages (i.e., pages that lacked information about their PhD projects), or whose pages had not been updated in five years or longer, were excluded from the study. Eventually, a list of potential informants was prepared, but it is worth noting that the choice of which potential participants were contacted was not totally pre-determined. Instead, the sampling process was iterative and evolutionary. For example, after the first round of interviews was conducted, initial analysis of these interviews determined those chosen in subsequent rounds. Participants were contacted and invited via email, with an invitation letter that included a short description of the study's purpose, as well as a consent form.

A profile of the study's participants is provided in Table 3.1.

Over the course of the study, a total of 26 search tasks performed by 9 participants out of 10 were observed during 12 hours of contextual interviews. Compared with quantitative studies,

the number of participants in this study is small, but qualitative research favors small samples, nested in their contexts and studied in-depth (Miles, Huberman, & Saldana, 2014). Moreover, small numbers are the norm for studies that primarily are exploratory in nature, i.e., investigating a relatively new phenomenon, in which data richness is of greater concern than the number of responses. Because of the nature of qualitative research, as Cleary et al. (2014) pointed out, inevitable challenges surface during various phases in all qualitative studies entailing issues such as how many people to interview and how one can be confident that no other important ideas will emerge. Despite all the debates around sample size, there has yet to be a precise answer to the question of what would be the ideal sample size in qualitative research (e.g., Marriam, 2015)? It all depends on the study, its goals, and the depth of the research, resource limitations, etc.

### 3.2.3 Instrument

The interview design was semi-structured and included open questions. Based on a review of extant literature and the researcher's background knowledge, interview questions were developed to encompass users' information needs in searching for cultural-heritage documentation. A semi-structured interview guideline was designed to collect data systematically (see Appendix A). The interview questions are grouped into two parts. The first part comprises introductory questions, asking the participant to describe briefly his or her research background, current research area, and the information sources and information systems they are using to search for their information needs. The second part of the interview encompasses participants' recent or current search experience and starts by asking the participant to perform a search task on his or her own based on the user's recent or current information needs. This question helps bring both the interviewer and interviewee into a rather real search-task situation for subsequent questions to elicit the participant's actual information needs.

## 3.3 Pilot study

Before the pilot study and before starting any data collection, pre-testing of the interview guide was done with two PhD candidates who were colleagues of the researcher. This pretest's purpose was to try out the general interview questions and see how the interviewer should approach participants to verify the questions' clarity, determine the feasibility of answering these questions, and build confidence around the interview guidelines. The pretesting of the interview guide indicated that it may be difficult for users to remember their recent search task(s) and their details in case they do not have any current information needs. Therefore, an email was sent to participants one week before their interviews, in which they were asked to provide some of the recent or current search terms that they have been using in their projects. This way, they would have search terms on record, making it easier for them to remember and replicate the same search tasks during interview sessions.

Data collection began with a pilot study conducted in March 2015. The researcher conducted

the pilot study after a pre-test of the interview guide to assess the research design's reliability, as well as ascertain the data collection's viability. The pilot study was conducted to test the interview guide with respect to formulation of the questions and to estimate the time required for the contextual-inquiry session. One PhD candidate who used archives and libraries was selected to participate in the pilot study. The contextual interview began by asking the participant to talk about her professional research background, current research project and related aspects, and the types of information sources and systems that she usually uses to meet her information needs in relation to her PhD research project. The interview was complemented with observations in which the participant was asked to conduct or replicate a couple of her search tasks triggered by her information needs. While observing the participant, the second set of semi-structured questions was asked. While performing the search tasks, she was asked, for example, to explain what she is looking for, what kind of information she needs, what she expects to find, and how she finds specific items relevant to her needs, what kind of information elements she uses to narrow the search or results list, along with anything else she would like to find through her query, regardless of information-system limitations. Finally, the contextual inquiry was concluded with a discussion about what the participant's ideal information system would be like.

The questions were open to provide the opportunity for more personal statements and reflections. Moreover, during some parts of the interview session, the researcher let the participant speak for herself in hopes of getting more useful information without the questions interfering with or guiding her answers. For example, she explains about a couple of information needs that she could not search for in online information systems and wound up finding them manually. The pilot study on the interview guide's research design was successful, with the collected data indicating no need for revision or modification of the interview guide. Since the data collected from this pilot study were significant, they were considered part of the principal study.

### 3.4 Data Collection

This study's data collection is qualitative in nature. To explore users' cultural-heritage information needs, having learned from the pilot study, a contextual inquiry included semi-structured, in-depth interviews, along with observations, chosen as the most appropriate method to collect data in a natural setting. This interview technique enables the researcher to go deep into a described experience, ask for more details, go back to an earlier point, request for clarification or explanation, and use observational and social skills for further discussion about an issue (Charmaz 2006). As this study has chosen an open approach to the subject, the framework, in connection with data collection, could not be too fixed. Therefore, it focused on performing semi-structured, in-depth, qualitative interviews with PhD candidates who were potential users of different memory institutions (archives, libraries, and museums), while observing them conducting their own searches. This study's observation method was unstructured and entailed studying participants' spontaneous behavior in natural surroundings to increase validity.

Potential participants were contacted through invitation emails including a consent form in which the study's purpose and the practical aspects of data gathering were explained. In a couple of cases in which the researcher did not hear anything back from participants, and there was a limited number of potential participants in a particular field, the researcher contacted them through their office phone numbers to ask whether they wanted to participate in the study. Nine out of 20 PhD candidates who were contacted agreed to participate via responses to researcher's invitation emails. After the pilot study, major data collection began in April 2015. Ten interviews were conducted, including the one in the pilot study, over four months. The researcher conducted the interviews either at participants' workplaces or in researcher's office, depending on participants' preferences. Most of the contextual interviews were performed at participants' workplaces. In three cases, in which participants did not have workplaces conducive to interview sessions, they preferred coming to the researcher's office.

Contextual interviews and observations – like those conducted during the pilot study – took place during each session. These contextual interview/observation sessions lasted between 45 and 110 minutes each, depending on each participant's time constraints and eagerness to talk and perform search tasks. All the contextual interview sessions were audio-recorded to ensure detailed data collection. Through the consent form attached to the invitation email, the participants already were informed that the interviews would be audio-recorded. Almost all participants (except one) replicated or conducted some search tasks regarding recent or current information needs related to their PhD projects. With respect to search tasks, unlike many other IR studies, we tried to keep the setting as natural as possible. Thus, we did not use any predefined search task, predefined information need, nor any predefined information system. The participants tried to satisfy their own information needs related to their PhD projects. They were free to perform their search tasks through any information systems they preferred or felt comfortable using. They also were free to conduct as many search tasks as they wanted using their own query formulations in the information systems that they chose.

### **3.5 Data Analysis**

Qualitative data, by its nature, according to Ritchie and Lewis (2003), invariably is rich, unwieldy, and unstructured, requiring that the researcher provide some coherence and structure to it while retaining the essence of the original data. The data collected for this study elicited a large amount of unstructured data, including audio files of interview sessions, observation notes, verbatim transcriptions of interviews (including different search tasks observed, along with interviews), memos (research journals), and snapshots from informants' replicated search tasks. The methods used for qualitative analysis need to facilitate the tasks of defining concepts; mapping the range, nature, and dynamics of phenomena; creating categorizations; finding associations; seeking explanations; and developing new ideas, theories, or strategies (Spencer & Ritchie 2002, p. 309). Using an inductive approach, grounded theory for data analysis inspired this study, but this does not mean necessarily that grounded theory was followed rigorously in every aspect. According

to Corbin and Strauss (1990, p. 6) grounded theory “has specific procedures for data collection and analysis, although there is flexibility and latitude within limits”. Grounded theory is a research approach in which researcher constantly interact with data using his or her own thinking strategies to make sense out of data (Corbin & Strauss 2008). In addition, inductive analysis, as the principal technique in grounded theory, means that the patterns, concepts, and categories of analysis derive directly from the data. While working with data, qualitative researcher develops sensitivity to what is in the data. Sensitivity, or insight into data, enables researcher to understand what is being described in the data and present participants’ stories with an equal mix of abstraction, detailed description, and feeling (Corbin & Strauss 2008).

The researcher performed analyses simultaneously during data collection, as is preferred in qualitative studies. A research journal has been kept, in which emerging analytic ideas and hypotheses have been noted, developed, and fed back into the data-collection process. As a result, data collection and analysis were an iterative process in which each round of analysis guided the purposive collection of more data by defining points of attention, the way the questions should be formulated to get better answers, areas that need more exploration, and participants who should be selected. Each round of interview sessions led to a better understanding of the phenomenon under study. These initial analyses were achieved based on the researcher’s memory, field notes, recorded interviews, and interview transcriptions, if they were ready. Simultaneous data collection and analysis are particularly important in inductive qualitative studies, in which, as Merriam and Tisdell (2015) also stated, although the researcher knows the problem and selects a purposeful sample in which to address the problem, he or she does not know what will be discovered, what or who to concentrate on, or what the final analysis will be like.

All 10 audio-recorded contextual interview sessions (i.e. 12 hours of interview) were transcribed verbatim – half by the researcher (interviewer) and half by a third person (because of time constraints) – on 210 single-spaced pages in a 12-point font. The researcher selectively checked the transcriptions by the third person for accuracy. Furthermore, out of 10 contextual interviews, one interview in which the participant did not perform any search tasks on any information systems has not been included explicitly in data analysis. The transcribed contextual interview sessions, each including performed search tasks, are the primary input for the analysis to explore different patterns in users’ information needs and translate them into the ontology (CIDOC CRM and its extensions).

According to Patton (2015, p. 522), in viewing each qualitative study as unique, “the analytical approach used will be unique”. Data analysis in this study includes five different phases. For each phase, a specific method or technique was adapted for this study to get the most from the data. Unlike other qualitative studies that use interview data, thematic analysis was not enough to answer the study’s research questions.

### 3.5.1 Phase 1: Thematic analysis and open coding

Thematic analysis is a common method in qualitative studies for identifying, analyzing, and reporting patterns (themes) within data. Through its theoretical freedom, thematic analysis, according to Braun and Clarke (2006), “provides a flexible and useful research tool, which can potentially provide a rich and detailed, yet complex, account of data.” Following the inductive approach of this research, the data have been coded using an “open coding approach”, in which the researcher begins the coding process without any pre-existing categories and creates codes as suggested by the data itself (Patton 2015, Braun & Clarke 2006). In this sense, this form of thematic analysis is data-driven (similar to grounded theory), but data are not coded in an epistemological vacuum. The process of coding itself can be achieved in several ways. The method adopted in this study included reading the paper-printed transcription texts, highlighting text passages in the document, reading the text for the second time and identifying a proper code for the highlighted passages in the margin. Since the coding process is more cyclical than linear (Saldaña 2015), it required going back and forth several times, refining codes, merging codes, creating new codes, etc. After open-coding the interview transcriptions to get a perspective on the coding made in the individual interviews, a summary of each interview’s coding was made. Each summary includes all the coding from each interview grouped under each participant, along with a small portion of respective coded text just for clarification of listed coding. This provided the researcher with an overall view of all coding. The importance of thematic analysis during this phase is to provide context and explanations as a kind of support for the next analytical phases.

### 3.5.2 Phase 2: Modeling the progression of information needs

After open-coding the data and getting a general view of the interview sessions, it was necessary to develop an appropriate tool to conduct the rest of the data analysis. The development of the information-needs progression model was the first step of this analytical phase, described below.

#### Phase 2.1: Developing the model

In the preliminary analytical phase, the researcher created a model to map out the progression of users’ information needs. The holistic model of information-needs progression (Fig. 3.1) presents the concept of information need as a whole and in interaction with information systems. The proposed model provides a holistic approach to studying information needs within a qualitative methodology context. The model is situated in the world of information and assumes that an information need and every other related aspect – from creating the need to it being answered as a whole process – take place within some context, which is specific to a particular individual in a particular situation and time, and not always simply ending when the relevant items are found.

The introduction of the model’s components shown in (Fig. 3.1) is as follows:

- User context: User context is a broad term and can comprise anything that relates to a user and influences him or her during the information-need flow. To date, several studies (Ingwersen

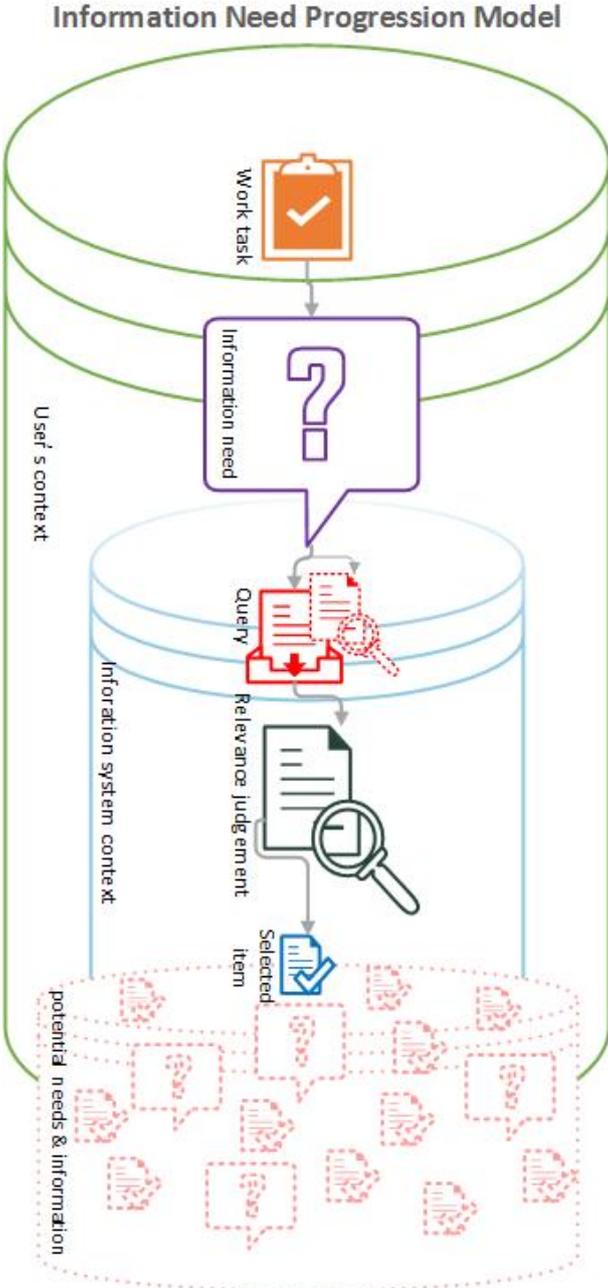


Figure 3.1: Holistic Model of Information-Needs Progression

1982, Pharo 2004, Marchionini et al. 1993, Marchionini 1990, Kuhlthau 1991, Nicolaisen 2009, Kuhlthau 1993b, Vakkari 2001) have highlighted factors that can be associated with the user context in this model. Different attributes that create a user's context include, but are not limited to, a user's traits (e.g., patience, tenacity, curiosity, interest, trust, ideology), skills and knowledge (e.g., professional background, knowledge of the field and work task, knowledge of the resources, knowledge of information systems, and search strategy), understandings and perceptions (e.g., mental models, the way the user perceives objects and environments), and overall circumstances (e.g., phase of research, time, work situation). User context always is present and provides context for other components of the model.

- Perceived work task: Work task refers to the task that leads to information needs' formation and respective search processes to be accomplished. Work task may seem not to originate in the user context, but it is defined and performed based on how a user perceives it, considering the user context (e.g., user's characteristics, knowledge, situation, time). The same work task can be realized differently by different users, and even by the same user at different times and in different situations.
- Information need: Information need is what the user needs answered. A work task or its sub-tasks trigger information need, but user context and, partly, the selected information-system context also can affect it. For example, a user's knowledge, situation, mental model, information systems, and their information coverage and indexing can influence the formation of an information need.
- Selected information system: A selected information system is a particular information system through which a user expects his or her information needs to be answered. User-context attributes affect selection of a particular information system, such as a user's perceived work task, his or her information need, mental model, knowledge of different information systems and their collections and coverage, personal situation, and time. On the other hand, the selected information system itself provides context for the model's other components. Information-system context includes information coverage, indexing, presentation, and all their constraints.
- Query: A query is what a user enters into the search box of an information system in hopes of finding desired information. Other components influence query formulation, such as information need, user context, and the information-system context.
- Relevance-judgment process: The relevance-judgment process is a process in which the user employs different strategies to evaluate search results. User context and information-system context can sway the relevance-judgment procedure. A user's characteristic, mental model, knowledge, situation, and information system's result presentation affect relevance judgment process.
- Selected item(s): A selected item refers to an information resource that the user selects from a search-results list as being relevant to his or her information need. The item that has been selected may satisfy a user's information need or may provide the user with some insights. What

the user selects as a relevant answer to a particular information need can be affected by the user's information need, characteristics, mental model, knowledge, situation, time, and/or the information system's presentation of results.

- Remaining or potential information needs and related information: After selecting relevant items that may seem to be the end of the process, another component entails remaining or potential information needs and related information, of which the user might not be aware. This component includes: 1) relevant information needs that the user is not aware of, i.e., visceral needs, according to Taylor (1962), or may realize are relevant through interactions with information systems and information found while performing a search task; 2) information needs that a user is aware of, but could not formulate or express in current information systems; 3) relevant information that users could not find in the current system for different reasons, including system and indexing-process shortcomings, as well as weaknesses in formulated queries or improper search strategies; and 4) other related information that exists in the bibliographic universe and can be found through other known/unknown information systems. This component shows that the information need and followed search session entail a continuing process. It seems that integrated systems, which contain well-organized semantic information and support exploratory searches, may hold the potential to help users find all their related information needs.

The holistic model of information needs progression is surrounded by two columns, including user's expressions on the left and the researcher's interpretation on the right side of the model. The two columns contain boxes associated with each component of the model and are differentiated by different colors.

Developing the information-needs progression model was a great help in analyzing data when undertaking a research goal that seeks answers to research questions.

### **Phase 2.2: Modeling the progression of information needs**

After developing the holistic model of information-needs progression, it was necessary to identify search-task segments within all interview-transcription text. Therefore, all the transcriptions were read one more time, but through a new lens. Considering the different components of the information-needs progression model developed for this study, this time, the intention of reading the transcriptions was to extract specific related data to different components of the model when users try to satisfy particular information needs performing search tasks. At this stage, research data were analyzed by identifying the different associated components of each information-needs progression. The progression models were created with Visio software for each search task observed because an information need triggered every search task. One to seven search tasks accompanied each interview session, except the one interview in which the participant did not perform any particular search tasks. It is worth noting that because we tried to create a natural setting, sometimes the participant performing a particular search task remembered something else and jumped to another search task without completing the previous one. Thus, these incomplete search tasks were excluded from episodes in that interview session's narrative.

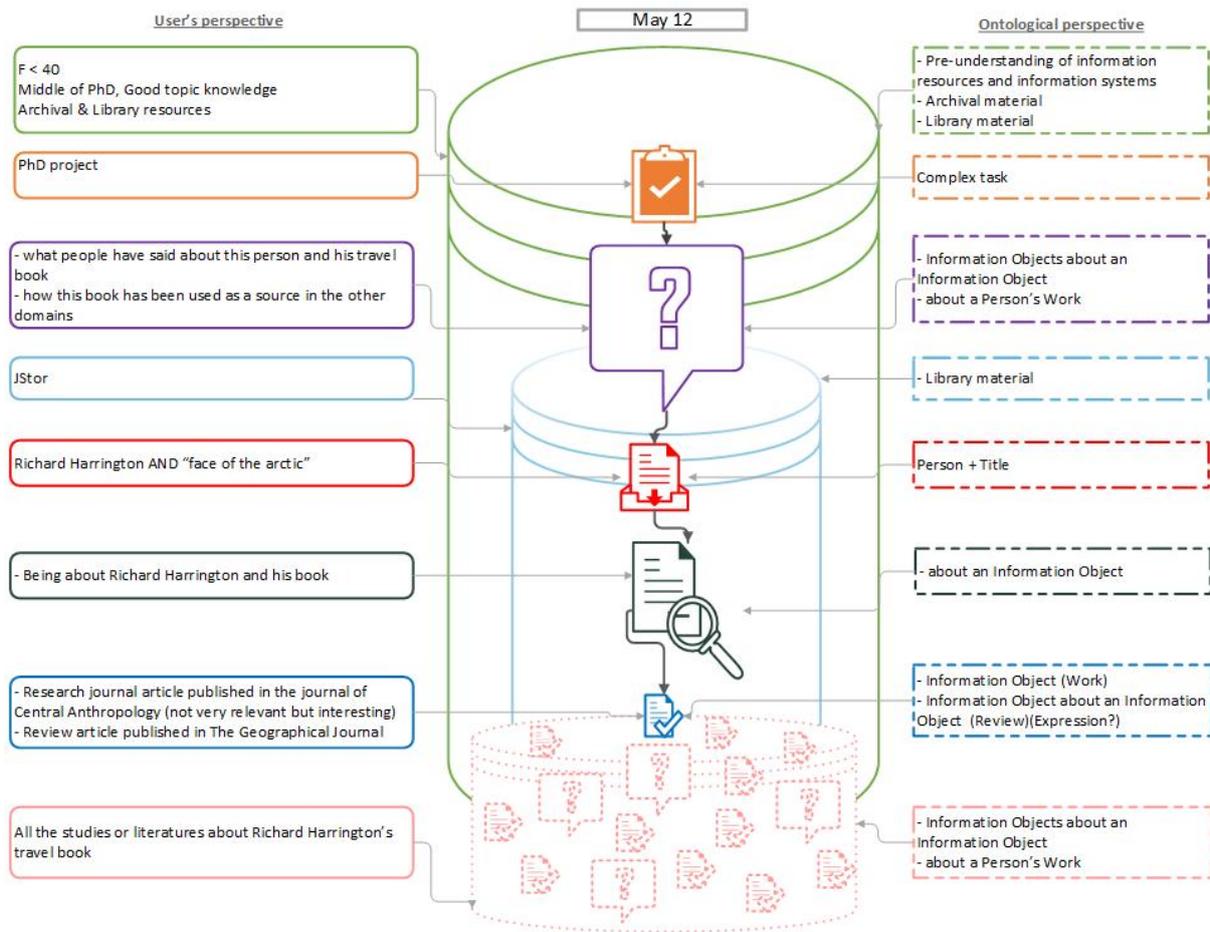


Figure 3.2: Sample of the Information Needs Progression Modeling

Fig. 3.2 shows an example of a created model. As can be seen in this figure, the information-needs progression model's various components are presented visually. On the left side, labeled user's perspective, related information on each component has been extracted from interview text and is based on the participant's expression. On the right side, labeled ontological perspective, a technical and ontological interpretation of the respective information on the left side is provided. The ontological perspective includes technical and ontological concepts such as attribute and relationship types.

### 3.5.3 Phase 3: Creating the stories

In this phase, for each contextual interview session, different expressed queries, along with contextual information related to the information-needs progression models, were collected and described through a story based on the interview-transcription text, screen shots, field notes, and the researcher's perception of observations. The stories provide not only a rich explanation of the abstract models, but also the context around each search session tied to an information need. These stories help illustrate the user's information need as a whole, which is important for further analysis. This holistic view, as opposed to a view of a particular part, provides a better understanding of users' information needs and decreases the possibility for bias from

isolating one part of an information-need flow. Each story starts with general information in an introduction, providing the user's background, his or her PhD project as a complex task, search behavior, information type sought, information systems that the user often uses, etc. It continues by describing different search tasks that the participant performed to satisfy information needs. Some verbatim quotes from participants are provided for better illustration and clarification. A story sample is provided below.

**Introduction** This participant is a female in her 30s and a PhD candidate in Social Sciences. She has a multidisciplinary background in Science (Environmental Science) and Social Sciences. Her project, as she expresses, is about cultural geography, urban geography, psycho geography, graffiti and street art.

The information resources that she uses are mostly books, academic articles, journal articles and her own photographs. She usually searches for information in Google Scholar (to look up specific articles), Google (for books), and Academia. "So mostly when I use Google Scholar, I'm almost always looking for journal articles. If I want a book or something I will just use Google". She believes that since she is doing an interdisciplinary research, she has to look for information in many different databases in order not to miss certain results. She explains that

my work is so interdisciplinary that it was very difficult, I would have to look in so many different databases [...] and my work is a little bit of this and a little bit of that. I would often be missing certain results.

[...] sometimes you don't want it to be too relevant or too restrictive, then you don't get, if you want to do interdisciplinary research, then you want to have access to sort of outside of your field. That's a problem I used to experience a lot with the databases, they were so restricted by discipline that if I was looking in a geographic one, I wouldn't have access to anything in art history or the art journals, which are very relevant for my work.

When she would like to do research on new topics, she would ask colleagues or people for their literature suggestions. It helps her to know where to start.

So that's often what I'll do, actually, is I'll talk to colleagues or people and ask for suggestions of literature, especially if it's something that I'm not that familiar with, the gender issues is something new to me and sometimes it's difficult to know where to start, which authors to look at. So often I will look for recommendations from other colleagues or friends.

When she gets the results list, she usually go through the first three to five or even ten pages to find the relevant results. However, she believes that "they put the most relevant, or the most cited anyways, first". To find the relevant results, this participant look at the journal titles because there are certain journals that are quite relevant to her field and she is looking for them. Moreover, she has a list of favorite authors who have similar research interests to her.

She is quite happy with the simplicity of results list in Google Scholar and thinks that too many categories in the results would be problematic. She prefers a simple result list that shows which one has the full text and so on.

### **Query 1: *Gender and power Connell***

**Information need: "an article by Connell on gender and power"**

In this example, the user is looking for a particular journal article. She knows the author and part of the title that the article is about it. Therefore, the user performs a query on Gender and power Connell that consists of part of the title and the name of the author on Google Scholar.

I was looking for an article by Connell on gender and power. It was a recommendation from somebody. So often it's kind of specific articles and an author that I know they've written and can't remember.

She finds the book that its full title is "Gender and power: Society, the person and sexual politics" by *R. W. Connell*. In this case, the user is looking for a specific known article by Connell. However, having found this particular article, she is still interested to see other works of the author on this topic. Because, she already knows that this author has written a lot on this topic. "He's written a lot on this topic so I was looking not for a particular one, just so see different things he had written".

In this instance, the participant is looking for a particular information resource.

**Query 2: *women right to the city***

**Information need: *specific article***

In this instance, the participant is looking for a particular journal article about the topic of "women right to the city". She inputs the terms in Google Scholar and browse the result list. A couple of titles catches her eyes, one of them is an article by a well-known author "Mitchell" that she is interested in his works and the other one is an article that she has read about it in another paper. Then, she finds an article entitled "The Right to the Gendered City: Different Formations of Belonging in Everyday Life" written by T. Fenster. that is exactly what she was looking for. To explore this topic more, she would read the references of the article, where she believes is "the best place to find the resources, especially if it's really, really relevant article, and well-written". She has found that this author (T. Fenster) has written a lot on this issue. She is interested to see all her works and having this intention, she goes to Academia to find all her works including published journal articles, not officially published articles, works in progress, conference papers, etc.

The participant in this instance is looking for information resources in particular.

### 3.5.4 Phase 4: Ontological analysis

An ontology often is illustrated as a set of concepts, as well as the relationships that connect these concepts. Using ontologies to describe and interpret users' information needs has been utilized in this study to get a more comprehensive understanding of information needs. This method allows for conceptualizing users' information needs semantically by identifying entities, relationships, and common patterns in users' information needs. Eventually, the emerging patterns can be used in ontology-based information systems to better meet users' information needs. Furthermore, it helps to see the possible gap between users' conceptualization of cultural-heritage information and domain-experts' conceptualization.

One of the best ways to represent ontological analysis is through network diagrams, defined by Saldaña (2014) as visual stories that relate research tales' major points. Saldaña says diagrams "serve the researcher by condensing the data corpus into at-a-glance references for reflection, and they serve the report's readers by enabling (an) at-a-glance cognitive grasp of the study's core findings" (Saldaña 2014, p. 147). Therefore, representing users' information needs based on the ontology (CIDOC CRM model and its extensions) through network diagrams would facilitate understanding by interpreting users' information needs through an ontological approach. Each of these diagrams represents a participant's information need and the query that he or she used

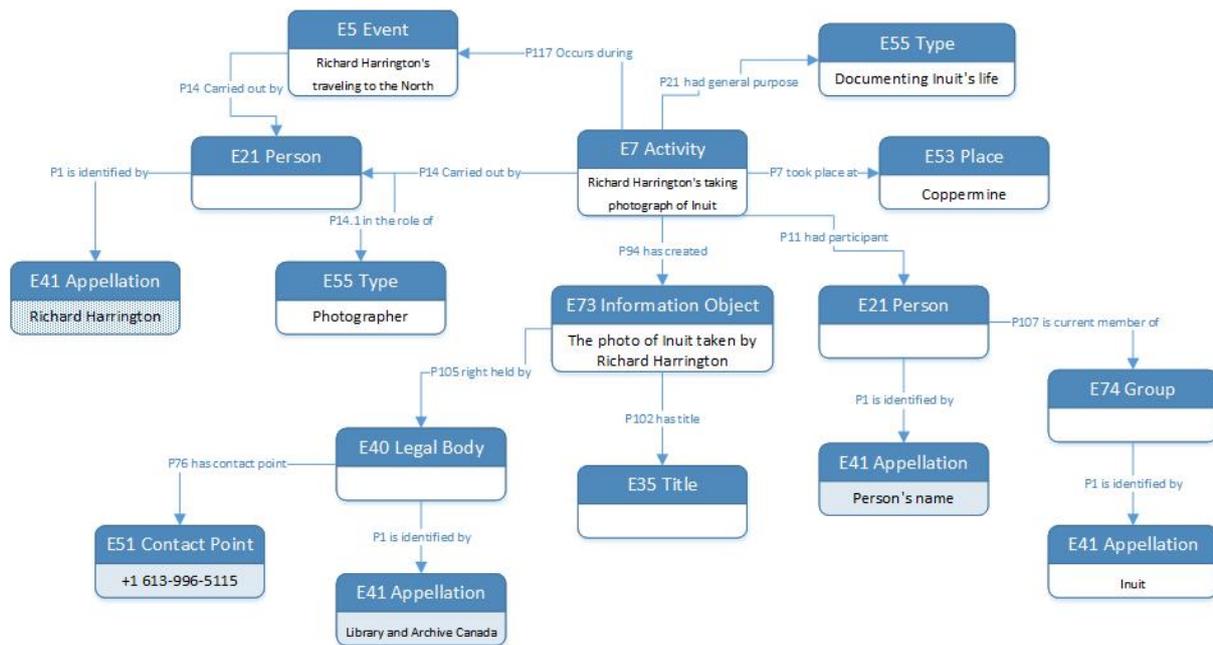


Figure 3.3: Sample of ontological analysis of an information need

in associated search tasks. Like the previous stage, these diagrams were created using Visio software. Fig.3.3 provides an example of network diagrams.

A brief description of this diagram is as follows:

In this example (Fig.3.3), participant inputs a query on *Richard Harrington* (a person's name) in the image search of an archival information system, which holds hundreds of the photos taken by this person in its collection. The participant's intention, using the person's name as the query, is firstly to find a specific photo taken by this person from the *Inuit* in *Copper mine*, which is one of the particular places this person visited during his travel to the North. However, what the participant actually needs to find through conducting this search task is the contact information of the right holder of that particular photo taken by Richard Harrington.

This diagram shows the relationship among all information entities in the information-needs progression modelling of this example and will be used for further interpretation and analysis. More explanation and interpretation of such diagrams have been provided in the Results and Discussion chapters.

### 3.5.5 Phase 5: Final analysis

Constant comparative analysis is done for each of the previous analytical phases for a deeper understanding of identify patterns, similarities, and differences. This analysis is achieved through comparison analysis on two levels: "within" cases (within each interview or each search task) and "cross" cases (across interviews and search tasks). A "within" case comparison, as Bazeley (2013) believes, leads to further relational analysis by sharpening awareness of case features, refining concepts and increasing sensitivity to contextual factors affecting the case. Cross-case comparisons are essential to exploratory studies that use inductive approaches. Memos that the

researcher writes during comparative analyses would be very helpful during the final analytical stage. In the end, the final analysis will entail the researcher writing an analysis that leads to what Creswell (2013) calls “substantive-level theory”, which emerges from the research memos during all stages of data analysis.

## 3.6 Limitations

The time required for data collection, analysis, and interpretation is lengthy. Because of the time and costs involved, qualitative designs generally do not draw samples from large-scale data sets. Therefore, general limitations in qualitative research are small sample sizes and large time commitments.

Because of the subjective nature of qualitative data and their origin in single contexts, it is difficult to apply conventional reliability and validity standards to it. For example, because of the central role that the researcher plays in data collection and analysis, it is impossible to replicate qualitative studies. Also, contexts, situations, events, conditions, and interactions cannot be replicated to any extent, nor can generalizations be made in a context wider than the one studied with any confidence. However, a qualitative researcher, as the primary instrument for data collection and analysis, is closer to the reality than any other instrument, with predefined factors and interpretation of reality obtained directly through observations and interviews. Thus, a qualitative study’s internal validity increases (Merriam & Tisdell 2015).

Some limitations exist that are specific to the present qualitative study. Audio-recording of the contextual interview sessions provided detailed data, but data analysis would have been less time-consuming if the sessions, particularly users’ search tasks, were video-recorded.

With respect to natural settings, it would have been more efficient if we could have conducted all interview sessions at participants’ workplaces (three contextual interviews were performed at researcher’s office). A couple of times, a participant could not perform specific search tasks because of not having access to a particular information system in his or her workplace, or else the participant could not remember the password to access a system.

## 3.7 Validity and reliability

Although it is difficult to replicate a qualitative study, we can facilitate reader generalizability by providing enough details about the study context (Merriam & Tisdell 2015). With respect to validity, the researcher have provided an in-depth and comprehensive description of the research design and analysis, participants’ settings, and findings through quotes. By embedding carefully chosen extracts from participants’ words, the researcher uniquely gave participants a voice in outcomes while contributing to the credibility and transparency of the research. The rich and thick description provided to contextualize the study boosts research results’ integrity through scrutiny so that the reader can determine the extent to which his or her situation matches the

research context and make decisions regarding transferability of findings (Lincoln & Guba 1985). Furthermore, purposive sampling was used in this study to maximize the range of information collected. The variation in sample selection allows for a greater application range for the findings.

The researcher increased reliability through consistent note taking and gathering of screenshots after the sessions. As a further means of ensuring reliability (or dependability in Lincoln and Guba's view), all the interview sessions were audio-recorded and transcribed verbatim, with data checked for any researcher effects during the observation phase of the sessions. Moreover, to ensure understanding, the researcher sought clarification throughout the interview process. These recorded materials, according to Lincoln and Guba (1985), provide a benchmark against which later data analysis and interpretations could be tested for adequacy and to enhance reliability. In other words, via a written research journal, field notes, screen shots, and electronically recorded materials, this audit trail created during the process of inductive analysis and theory development is both visible and verifiable. Moreover, to increase research design and instrument reliability, the researcher pretested both by conducting a pilot study.

### **3.8 Researcher position**

First and foremost, I am a researcher with a passion for providing users integrated access to cultural heritage information preserved in memory institutions (archives, libraries, and museums). I am interested in CIDOC CRM- as a core ontology in cultural heritage domain that facilitates integrated access to semantic information- and, in particular, in investigating how ontology-based information systems should present semantic information to users to satisfy their information needs and simultaneously not to make them overwhelmed with the complexity of ontologies. Regarding my background as a professional system librarian, I have worked with organization and retrieval of cultural heritage documentation for several years and experienced the growing need for integrated access to cultural heritage information. This double role helps give me insight into the subject of research. However, I intentionally selected people of different academic backgrounds from my own in order to counter any bias so that my understandings and experience do not shape the research.

### **3.9 Ethics approval**

Through the consent form sent to the study's participants along with invitation emails, the researcher informed participants of the nature of the study and that the researcher will not use participants' names in publications. Moreover, Oslo Metropolitan University and Norwegian Centre for Research Data (NSD) as the data-protection official for Norwegian research institutions approved the study's procedures and overall research project.

Chapter 4

# Results

In this section we illustrate participants' information needs in the form of stories as the result of the first two phases of data analysis- Phase 1: Thematic analysis and open coding and Phase 2: Modeling the progression of information needs)- on a total of 43 information needs expressed by 9 participants during 12 hours of contextual interviews. Each story begins with an introduction containing general information about participant's background and his/her information behaviour. It continues by different search episodes that were extracted from the interview sessions. The search tasks that did not complete were excluded. In cases where the participant performed too similar search tasks, only one has been included in the stories. Moreover, some information needs that could not be searched in information systems are also included in the search stories. The stories including both the introduction and search episodes provide the context that helps get a better understanding of information needs of participants in this study.

## **4.1 Story 1: User is interested to know about a particular fashion designer**

### **4.1.1 Introduction**

This is about a participant who is a female in her 30s. She is a second- year PhD candidate in history and has accumulated background knowledge of the target domain from her five years of research. Her current project has two perspectives: "business history and history of art". She said she would like to study not only the creations of designers but also the organization behind the fashion and style.

like how it is produced, where[it is produced], who is producing it, and how is the company organized? and so on. So, I try to have like a very round perspective around the companies. Not to just focus on the creations by the designer.

This participant has an interdisciplinary research. "I have a background of history, but I am looking at anthropology, I am looking at sociology. Every kind of social sciences." She explains that she uses of archives, libraries and museums and consults a variety of cultural information resources.

[...] I identify archival resources and physical objects, but also secondary resources, like newspapers, all kind of information I can find like internet and whatever. [...] It can multiply the resources to infinite level. Because I could interview with people, I could also brows for oral archive, video archives, like paper archives, but also books, ... literature, objects but also newspaper, all kind of items.

The information systems that she uses to look for information depend on on what in particular she is seeking. They include WorldCAT, Europeana, French National Library, and others. She

also has an understanding of the content of each information system, what she can expect to get from each information system or what to search for in each one. Such understanding comes from her mental models formed via interactions with different information systems. For example she notes:

“I would search more by designer here (Europeana fashion) . . . but definitely it would be more efficient if I search the designer name in this database [ . . . ] but this is a French library. So, it is more relevant to find book in French on French topics and I would brows WORLDCAT in English or the Library of Congress in English”.

This user is aware of having two different search strategies : exploratory and known-item search. Generally, she performs exploratory search:

[ . . . ] you never know what’s gonna be really interesting into your research. [ . . . ] I start with an idea that I would like to research . . . . at the end of your research, it’s like it does nothing to do with the beginning. It is a total different research

However, she expresses that when she would like to find something in particular she searches differently than when she is looking for new ideas. “So, if you expect the result or if you don’t expect the result is not the same type of research I do.” The participant makes an example for her two different types of search strategies (although in both she has kind of exploratory behavior):

If I expect the result like I know that Schiaparelli has produced this sweater late 20th, I would say and I know that it is start of her carrier. So, I like to look for everything I can in this is period. I am going to try the newspaper I know of, I am going to try the museum that hold this. I am going to try like everything we have done, we have talked about this sweater. But let’s say I want just to look for other idea about it. I probably try to find all the designers, maybe producers, maybe techniques. . . . I am going to type like random keywords and associated keywords.

She explained she would like to be able to rapidly identify the resource type in order to select the best result of the set.

[ . . . ] Is it type of publication, is it a book, a newspaper, is it a song, is it a picture? This is like to very quickly and easily way to identify what you are searching and then, you have the details if you just click on to it.

[ . . . ] Of course because if it is a book or a newspaper you are not definitely going to look at it on the same way. You need to know it first.

For the results, this participant prefers the list-based display because she finds it clearer and easier to understand. “[. . .] I really like the list presentation. I think it is much more clear. It is like Google presentation and it is like Google results and it is much easier.”

She does not like the pre-categorized results, in which you need to guess which category may contain your relevant results. She finds pre-categorized unpleasant because she believes that it is difficult to understand the logic behind the categories.

I like the list better actually, but what I hate I can say what I hate actually. This is . . . Library. And it has a great collection but result is still messy when you just do something [Searching. . .]. Oh, what is this? It is just like visually, it is crappy. You don't understand what you are looking for. It is like two lists and you don't understand how they sorted them and why is it. . . “Catalog” is separated from “articles”, from “academic commons”. It does not make sense to me. What I like is having a list [. . .].

In her ideal information system, she explained that one of the most important things would be “the capacity to make links with other information”. The links would help her go further and find the related information quicker.

[. . .] like leather (in the description of an object). If we click on leather and you see like other items in leather or other research associated. Like everything that make it quicker.

[. . .] It is really a huge gain of time; from a book I can go to the article, I can go to the garment, I can link to something else. It would be so much better.

She also likes sorting tools or filtering options based that allow users to select different criteria such as type of document, type of object, date of publication or creation, provider, subject, author and so on. She believes

“as longest series of criteria is better”. Such filtering choices helps her to refine the results and reduce the number of hits while having the whole view of the results as well. “[. . .] a way of having everything but being able to make it . . . into small packages”. It is pleasant for her that she can choose among the different filtering options herself. This is what she calls “a good way of putting your own query into the results.”

[In a desired integrated system][. . .] I would do kind of a quick search. I need a way to sort them very quickly, by type of items, by date of publication or creation, by provider, whatever. A lonely soft criteria which allows me to reduce the number of hints quite quickly. Because it can be like really intimidating if you have ten thousands results and if I say I am going to brows, today it is not possible.

In general, she is not very happy with the online search because it does not contextualize results.

[. . .] because I think what is missing for me when I do online searches is like you just have the exact results you wanted but then you miss everything around. It is really hard to get an idea. So, it is good if I can still have this paper-like way of doing it.

This participant thinks that both online search and the traditional way of finding information by going to the physical library or archive and manually seeking information are useful in different ways and complement each other. She believes that for object-based research, the online tools are sufficient to begin a search. However, consulting the librarian or curator at the physical museum or archive proves more helpful not only to get the information object but also to see it in the context of a collection. She explains:

[. . .] it [online search] is kind of easier and more comfortable also. and also you can do your own research and try keywords you think, you can dig, you can bump to another keyword, you can change but it doesn't replace the advice from librarian and curator because they know their collection. So they can put you on different tracks that you have not thought about.

Her different search episodes are as follows.

#### 4.1.2 Query 1.1: “*Intellectual property fashion*”

**Information need:** “*What is the research context around [intellectual property and fashion]?*”

In this instance, the participant is going to write a conference paper on the topic “*intellectual property [relating to] fashion*”. She would like to see what others have done on this topic and what the research context around this topic is. “what historians or sociologists have already done with this material? Or what is the context, the research context around it?” To find an answer to her information need, she enters keywords like “intellectual property fashion” into a search form in Worldcat. She explains this is an open query intended to generate and gather ideas. “It is really a first search. So, at this point I don't really know what I want to find. So, it's more to give me ideas.”

Having scanned a couple of result pages, she selects a book entitled “Innovative Design Protection and Piracy Prevention Act : hearing before the Subcommittee on Intellectual Property, Competition, and the Internet of the Committee on the Judiciary, House of Representatives, One Hundred Twelfth Congress, first session, on H.R. 2511, July 15, 2011” because she already knew about this book and general content. “it is basically all the debates around the intellectual property”. She has selected this book but as she mentioned in the beginning, works authored by

historians and sociologists could also be useful information objects that serve her information need. In this search episode, the participant is looking for information resources that are about a particular concept.

#### **4.1.3 Query 1.2: *History of property/ histoire de la propriété intellectuelle***

##### **Information need: “theoretical background”**

In this instance, the participant tries other keywords to find some information for the theoretical background of her conference paper. “I already know where the primary sources are and these are keywords for the secondary resources and the theoretical background”. Therefore, she searches for the topic “History of property” on the library search database of National Library of France. Then she performs a search in French with the phrase “histoire de la propriété intellectuelle” because she believes the results might be different. She chooses a book from the result list from the French language search. Because she would like to get some more information about the author of the book as well, she clicks on the author link and finds some general information about the author, including nationality, languages which they speak, and date of birth. Beside the general information there is a link to “other related records” Through related records, the participant can find both what the author has written, and what people have written about the author. Thus, this user has learned that in the bibliographic world, a person can be indexed both as an author and as a subject of an information resource.

Because with different people you can have author, subject, . . . for instance if you look for like Peggy Guggenheim; she was a collector and people have written biographies of her but she also wrote her autobiography. So if you want to see what she has written or what people have written about her.

Here, the participant successfully satisfied an information need. She began her search with a topic, found a desired author, then continued her exploration by browsing the author’s other works and the works about the author. Information resources are type of information that this participant is looking for in this instance.

#### **4.1.4 Query 1.3: *Elsa Schiaparelli***

##### **Information need: *find “the visual [representation] of the garment that is counterfeited”***

This instance again relates to intellectual property. The participant has prior awareness of a counterfeiting trial, in which a counterfeited sweater has been discussed. She seeks a picture of that original item.

I am working on intellectual properties and I find that there is counterfeiting trial. I know that the garments counterfeited are from 2013. Then I can browse this database to find the visual of the garments counterfeited which I won't have with the transcription of the trial. So, this is kind of to illustrate or to find like the visual part of my research.

In order to find the picture of this garment, she begins a search task on the Europeana Fashion database. The user searches for the designer's name "Elsa Schiaparelli".

Some results are Schiaparelli's own designs while others are designed inspired by her. Therefore, she tries to narrow the results selecting the "designer" as "persons' role". This tactic did not work, because Elsa Schiaparelli is indexed not as 'designer' but as 'creator'). She narrows the search once again using this technique. This time, she selects "knitting" as 'technique' and finds the sweater she was looking for. Since the participant's information need was to find the picture of this garment as information resource, in this case her search is successful.

**Information need: *Contact information of the information provider***

When this participant finds the picture of the sweater, she is interested in finding its original provider (the museum or institution), where this garment is held—and contacting that provider directly. She does not believe the Europeana Fashion database alone has sufficient tools or information to satisfy this need.

I think this portal is good to give ideas of who I should contact. But I wouldn't trust in their information and I would still go to the original provider to be sure I am not missing something.

She suggests she looks at the portals like Europeana as a starting point that leads her to the people and museums she should contact to get more information.

"if I know the provider then I can contact them directly for asking relevant question. But just information itself is not sufficient. I mean the information of the garment". After the participant finds the main provider "Museum of Decorative Arts" where the sweater is kept, she goes to their website, checks their online collection and finds their contact information. She suggests she will contact them and ask if they have any other designs of Schiaparelli, if they have any other items related to this garment, among other things.

Through browsing the results, she also finds a new institution (Museum of Applied Art in Belgrade), that also has one of Schiaparelli's designs. She was unaware of this place and holding and is excited that she happened on this particular result.

[Browsing via other successful results sets] is a good way of finding providers [you otherwise would not have know about]. for instance I would have never thought

about checking the Museum of Applied Art in Belgrade. I have no reason to go to this museum but then I can say OK that is interesting.

[...] well, they don't have online collection... but I might send a note to the curator and say like I have seen in Europeana that you have a shoe of Schiaparelli. I want to know if you have any other items for instance.

For this participant, the provider or the place where the information is kept is very important. Because she can contact them and get more relevant information from them. In this instance, the contact information of the original provider including email or telephone number that is kind of fact-based information is what this participant would like to find.

### **Information need: “*what is the quality of the textile?*”**

In this instance the participant has not been able to fulfill the information need. Pointing to one of the objects in the results, she says:

But even though if it says like the material is leather, but that is all it says about it which is obviously not only leather because you have all these decorations which are not in leather. And it just says shoes. It does not say like, because all shoes can have a name because of the shape of the shoe. So it does not tell you this is like a Salome or whatever. It just tells you it is a shoe [...] when you do an object-based research, it is really important you see the quality, you see the sewing, you see everything. If it is hand sewn, machine sewn, what is the construction of the garment, what is the quality of the textile. So, it is really important to have this kind of real object dimension. So, this is a knitted sweater really famous. They have put like the dimensions, black and white (color) and wool... It gives you a kind a good idea of what item is. Then I wanted to visit the collection and I was seeing it kind of really rough wool. Really tough and really delicate. And this's something that picture will never give you.

In this instance, the participant is looking for more detailed facts about the museum object and available metadata was not sufficient to engage the object in the way that was most important and valuable to her

### **Information need: “*how many items have survived?*”**

This participant would like to see all the other designs by by a particular designer—in this case, Schiaparelli— and additional information including the designs' locations.

Resp.: [...] I might send a note to the curator and say I have seen in Europeana that you have a shoe of Schiaparelli. I want to know if you have any other items for instance.

Q: So, you would like to see whatever she has designed.

Resp.: Yes. and then, if I have the chance like because what is interesting is to know where these items are held and how many items have survived, how many dresses, how many accessories and then if I have the chance, I try to go and see them.

In this case, the participant is looking for all the designs by Schiaparelli that are considered as information-resource finding. Moreover, the information resources should be relatively metadata-rich in order to satisfy the information need.

**Information need: “*how many museums have a sample of this sweater?*”**

She would also like to see how many samples of a garment—in this case, a sweater— have remained and how many can be found at the museums. She explains this gives her insight about a garment’s relative popularity, which to her is an important attribute: “the more pieces you can find,” she says, “the more famous they were”.

In this instance, the participant is actually looking for the context around this sweater —not just information to locate it.

**Information need: “*in the industrial side, who was producing the wool?*”**

The participant would be also interested to know other aspects and procedures related to this sweater. Notably, she found this kind of information previously with the help of a curator—not by interfacing with a digital system.

the curator [...] told me that she has a file that she uses for the exhibition and she had like a research file for knitting sweater. . . she had the research and in the industrial side who was producing the wool. . . I would have thought like who was designing this sweater but not who is producing wool. It just other insights and other perspectives. This is something you can’t really do I guess with just the browsing tool. . .

In this instance also the participant is looking for some context around the sweater —but needs human assistance, she says, to fulfill the need in question.

**Information need: “*the story of this garment*”**

The user says she is interested in more information about the garment, saying she wants its story or “context”. In this case, she wants to delve deeper into the narrative history of its design.

Resp.: [...] But just information itself is not sufficient. I mean the information on the garment.

Q: So, the related things would be more interesting?

Resp.: yes, the context again. I think it is quite important.

Q: for example for this garment, what are the things that you can say these are the context?

Resp.: what is interesting with this specific garment is that she . . . , I mean . . . , the story of this garment is that that she [Schiaparalli] has started her carrier by doing that [. . .]. It was really casual fashion. You could use it to play tennis, golf, whatever. It is really particular. Because it is like hand sewn knitted. It is a knitted adult jumper. And it is kind of really rough design. It is far from being perfect [. . .] It was this kind of rough aspect that she was looking for [. . .] It is quite like humoristic or funny design and it was really different from all the other jumper. Other jumpers were really perfectly done with really sophisticated designs. But this one is rough and imperfect and really easily wearable to do sports and lets you to have movement.

What this participant would like to find in this case is again some contextual information around the sweater.

**Information need: “*similar design of another designer*”**

She would also like to see the samples of sweater that have been designed by another designer in order to compare the differences between Schiaparelli’s design and her another contemporary designer’s design.

In this instance although she is going to find another design that is a type of information resource, she is looking for some context around the sweater and its designing.

**Information need: “*if knitting [a] sweater was [. . .] common*”**

In addition, she would like to know more about “the knitting market” around the time Schiaparelli designed this garment.

And I would like to know more about like the knitting market like who, I mean in 1927 I guess like every woman could knit. But I would like to have information about if this is like my own perception and if it is true or not. If the knitting sweater were like a common item for instance; if it was only for sports if it was only for winter. . . I mean all these kinds of information behind

The user again needs contextual information to satisfy the information need (and build up more narrative history around the item).

**Information need: “if she was becoming famous with it [the sweater]”?**

The participant is also eager to see if this particular garment has been represented in any other publications. If so, she wants to see which newspaper or magazine published a picture of this particular sweater, and the time it appeared. She would like to know if the newspaper or magazine used a picture to advertise or if a picture appeared in a story signaling Schiaparelli’s rising star. She explains:

I would like to look into the newspaper and to really know if she was becoming famous with it or when her name started to appear, when was the first time that the newspaper wrote about her. You know this kind of information. What was the fashion journalist saying about her. what was the reaction from the public. I just like the contemporary reaction about it

In this instance, the participant is again needing “context”.

#### **4.1.5 Query 1.4: Schiaparelli**

**Information need: “What did people think about fashion at that time?”**

This participant is doing an exploratory search and consequently she has many different information needs that are related to each other and allow you to follow the breadcrumbs. As she says “. . . you just start with the garment but you can spread like. . .”

Following up her previous query, in which the participant has found one of the designs of Schiaparelli, she performs a new search task in the newspaper and journal collection of National Library of France. She would like to find some more information about the garment and the context around its designing. Therefore, she browses the newspaper and journal collection and then queries on the name of the designer (Schiaparelli).

In this instance, the participant first finds a particular known magazine (Femina) through browsing based on type of document (newspapers and journals) and the topic (feminine), then searches for the designer (Schiaparelli) within the browsed results. She then continues again with browsing based on the date. She already knows the approximate period of time when the garment was designed.

She finds Schiaparelli and one of her designs in an issue of the magazine. However, she still needs more information situating Schiaparelli and her designs within her contemporaneous social and consumer context. When the participant is asked to explain what else would be interesting for her to give her the context, she articulates several needs. For example, she would like to know about the fashion and preoccupation at the time of designing the garment both among the people who are the consumers and among the other designers who were Schiaparelli’s contemporaries. How different or similar were Schiaparelli’s designs to other designs of her time, and what makes

Schiaparelli's designs so significant in comparison?. In order to find this kind of information the participant would like to see not only the page about Schiaparelli's design but also all the other pages in that issue of the magazine— even the pages that are not about Schiaparelli and her designs.

I have [found] my [Schiaparelli's] garment [in this journal]. I would like to browse it quickly the pages around it [. . .]. What I know is that Schiaparelli has created this special coat or dress but I do not know what was fashion in October in this month. Was she the only one to present this type of garment? Or was it really trendy. Everyone was doing it? What was topic of preoccupation like? What did people think about fashion at that time? What was their interest?

For instance, I find that in this one, there is fur and she was not using fur in her dress. Let's see if fur was important at that time. And here is an article about [winter . . .]. So, maybe if I read it I am gonna get information about what she created. Even though it is not exactly about her. So, this is what I meant by context. It is really important but I can't quantify or qualify it. Because it depends really on what I am searching.

#### 4.1.6 Query 1.5: *Elsa Schiaparelli*

##### Information need: *“just fishing for ideas”*

This participant already knows about “Elsa Schiaparelli” as the person who designed the sweater of interest at the start of her fashion career. Now, however, her needs are not specific; rather she is “fishing for ideas”. She does not even expect particular results. This means that this user does not look for a particular type of resource or specific object, and instead puts a query with the designer's name “Elsa Schiaparelli” into the general search box of National Library of France to see what she can find there. The first item in the result list that catches her attention is a letter that Elsa Schiaparelli has written an individual named “Christian Bérard”. This letter seems to the participant an interesting, perhaps more personal, finding, from which she can continue her exploration about the designer.

The user describes her exploratory search as kind of navigation from one topic to another to get an idea. Her preferred way of navigation would be the possibility of having hyperlinks and following them. In this example, after finding this letter and the addressee, she gets an idea and she becomes interested in finding information about Schiaparelli's professional network. She would like to know with whom Schiaparelli has collaborated, as well as information regarding which projects were collaborative has been collaborating during her professional job and what have been those collaborations on. In this instance the user seeks inspiration for new research angles through serendipity. What the participant is looking for in this instance, is some contextual information to give her ideas and provide her better understanding.

**Information need: “Who is Christian Bérard?”**

After that the participant sees some descriptive information of the letter, she is curious to know who Christian Bérard is. She would like to find more information about him and his type of collaboration with Schiaparelli. Therefore, the participant clicks on the name of Christian Bérard to see who he is. The link leads her to a page that shows some information such as his nationality, date of birth and death, and his role based on the library’s available documents. From this page, the participant learns that Christian Bérard is an illustrator. She still needs to find more information about this illustrator.

Here, the participant is looking for all related information about Christian Bérard as an entity.

**4.1.7 Query 1.6: Christian Bérard**

**Information need: “If he was a famous, if he was a minor person”**

The participant continues her exploration with a new search on “Christian Bérard” on the same database. The results show a collection of documents that she decides to refine it by documentation type as books. Browsing some items, the participant does not look very satisfied with the results. The user still wants information about this person. So, she tries to find more information somewhere else through other information systems and databases.

In this instance, the participant is looking for some context.

**4.1.8 Query 1.7: Christian Bérard**

**Information need: “What he looks like?”**

Following up her previous information need, the participant is interested to see what Christian Bérard looks like. Therefore, she just searches the name of the illustrator “Christian Bérard” on Google and goes to Google’s Image to see pictures of him.

**Information need: “If he was a famous, if he was a minor person”**

The participant still would like to know more about Christian Bérard and his collaboration with Schiaparelli. So, she gets back to the result list of her Google search and finds Wikipedia page on Christian Bérard. She reads through it and follows the related links in the page such as his father, his influences and so on to see what she can find that interests her.

you start to be in Wikipedia and then you just gonna click on his father and from his father I am going to go back and there is his influences. Oh wow it is interesting. And notice a painter. . . (one of his influences is a painter).

The participant would like to know if he has had an important collaboration with her. She is also interested to know about Christian Bérard's collaboration with other designers in the same period to see how popular he has been.

[...] about how illustrator and her have been working in this period. If he was a famous, if he was a minor person, then I am going to go back to Schiaparelli to see if the other artists she has collaborated with. And if he is important in her collaboration or not.

In this instance, the participant first look for all the related information about Christian Bérard as an entity. Then, her next information need is to find some context around Schiaparelli's professional network.

## **4.2 Story 2: User would like to find information about a particular law**

### **4.2.1 Introduction**

This story is about a participant who is conducting a research on human rights and ecology. His PhD project is in a multidisciplinary program of law, history, politics, and social sciences. He has done different general research on human rights and ecology. However, this participant sees himself new to the specific case, here a particular law, which he is going to do research on during his PhD program. The information systems he uses to look for information includes Oria local library system, WorldCat, and Google.

Regarding the information resources and materials, this participant explains that he uses different kinds of resources such as books, journal articles, government reports, newspaper articles, commentary articles, news reports, audits, evaluation of the acts, videos and legal documents.

The participant emphasized that he does not like to be too specific in search for the fear of missing some information. He prefers to use broader terms as his query and browse the results himself to find the related things.

I don't like to be too specific with searches because I find that then I miss things. I'd rather spend a lot longer, you know, if 500 had come up I would have gone through all 500, maybe only 5 are relevant but at least then you know you've done the search. So that's kind of how I approach things, I like to be very, very thorough in that sense. So that's why I try to use quite broad terms at least at the start or when I'm doing something.

In the early stages of his research, he usually performs recall-oriented searches and begins his search with a general query to retrieve as much relevant information as possible. He does such kind of broad searching specially in the early stages of his research and narrows it down later as he mentions “I wanted to be as open as possible at first at least and I’m like then narrow it down a little bit more”.

When it comes to filtering the results, he does not feel comfortable to exclude some categories of the results. For example domain or subject categorization seems confusing to him. He is a little skeptical about filtering the result through subject categories because resources can be indexed under different subjects by different people. Therefore this type of filtering is worrying to him and he has the concern of missing something. Several times during the interview he expresses his concern about missing some relevant information.

So I’d rather not use that to limit it because again I want to be, I don’t want it to be based on other people’s categories, I want to have it at this stage some control over how the information is categorized.

I find like the way the subjects are organized sometimes I’m a bit confused by. I mean what people mean by ecology or human rights again can be so different. So again I don’t want to exclude stuff because I’m a bit worried that I’ll miss something. But yeah I probably could exclude for example social epidemiology, I know I’m not interested in epidemiology or I have no way of really approaching that.

Regarding language filtering, he believes that even though he can not read books that are in Spanish for instance, just knowing that there are some related books written in Spanish is still important. He says: “I get nervous about excluding, I’d rather have like a big search and only be one thing and not to miss it. Even knowing that there’s stuff that’s been written in Spanish is useful for me”.

To see if the results are relevant to his research, he goes through each result and would search for the topic inside each book or article (if applicable) to see where his searched topic is located. If he finds it in the footnote, it would not be relevant. This participant does not like neither to follow the recommendation of the system regarding the related resources nor to follow the bibliography list of the articles to find his relevant information resources. He explains that:

So I’d much rather spend a whole day going through the entire search terms than starting with one and then going to the one that’s most relevant to that and so on and so forth. Because I kind of, like I want to have some control over what is [relevant or] what I consider relevant not what this system says relevant or what the authors themselves consider relevant.

I am aware just as these systems [like Google] in a sense filter and prioritize that and what I'm getting is slightly filtered, so that makes me also again very aware of [that] I need to go through it all myself and make sure I haven't missed anything and categorize it myself in a sense.

The participant has some information needs and performs some search tasks on information systems that are described as different episodes as follows.

#### **4.2.2 Query 2.1: *Rural employment guarantee***

##### **Information need: *Everything about and around the Mahatma Gandhi Rural Employment Guarantee Act***

In this instance, the participant is searching for a particular Indian act that is called Mahatma Gandhi Rural Employment Guarantee Act. It is not a general topic, but a specific topic, which is a particular act. The participant performs a search task on Google and the query that he uses is the popular name of the act that is "Rural employment guarantee". Through this query, he needs to find all the information about this particular act. He would like to know everything about the act, from its formation and the people involved in to its practical implementation, its evaluation and the way people perceived it.

... because I'm interested in how people perceive the act as well, not just, so if you like the official information it's very useful for judging how the act is practicing one way but I'm also more interested in how the people involved in both who are generally the participants in the act but also people who are interested in the act and when did the act come about and who are involved in forming it and involved in evaluating it, how they perceive how it's worked in practice. So this whole media for example can be very important in giving a taste of what's going on. (p. 5)

Besides, this participant explained that she is also interested in other aspects of this act like the gendered aspect of the act. He encounters a book entitled "The Gendered Impacts of Liberalization: Towards Embedded Liberalism?" that seems interesting to him. Although the name of the act has not come neither in the title of the book, nor the abstract or introduction, one of the sections is about this act.

This participant in this instance is looking for all related information about this particular act as an entity and the context around it as well.

### 4.2.3 Query 2.2: “*human right*” and ecology

#### Information need: “*who is writing about this?*”

In this instance, the participant who is at the early stages of his research project performs a search task in a local library system called Oria. The query that he uses is “human right” and ecology, which is a topic. The information need behind this topical search is quite different from the other exploratory topical searches, in which users are looking for information or resources that are about the topic. The intention behind searching for this topic, besides finding related books or articles is to find the people or related organization who have written about this topic. It is because he wants to find his potential interviewees. “. . . well the information I’m looking for is not just the book or the article, or the actual text, I’m also interested in who is writing about this”. Here the participant is looking for the names of people who worked on his desired concept that are fact-based information type.

#### Information need: “*how many times they’re writing about the same issues?*”

This participant is also interested in the authorship context around the particular topic. For example, he is interested to know how many times particular co-authors have been writing together on specific aspect of a topic.

But it’s interesting to see how, I mean there was a group of authors, three of them, who often co-write stuff together. . . But it’s interesting to see how many times they’re writing about the same issues. So that kind of people I might want to talk to about the law in more detail, like actually interview.

So I might not necessarily read everything that they’ve written but just to know that they’ve written ten articles on virtually the same aspect, this law is very useful information to get out of these searches.

Here, people who have worked on this topic several times would be the ones who this participant is interested in talking with to get more detailed information. In fact, in some cases finding these people are more important for him than what they have written.

In this instance, the participant is looking for some contextual information to get an idea of who he should interview with.

## 4.3 Story 3: User is looking for information about social system

### 4.3.1 Introduction

The participant is a female in her 50s and a final-year archaeology PhD candidate with 30 years of professional experience as an archaeologist. Her PhD project is about colonization after the last Ice Age. She is studying the colonization process and development, the craft tradition of that time, the type of contact patterns throughout it, the stone tools used and their production, and other overall traditions in detail. “[...]it’s a very detailed study of this craft because the craft of making tools for example, they were thought from person to person you know, between individuals”. “So you have to understand the same as you can see the difference in the pottery making, not only the patterning like decoration and the forms of the pottery, but also the shapes, but also how they actually made it you know, that is step by step how they actually made it, this is a very cultural thing, yeah. So that is what I’m studying in the stone tool production, yeah”

She couples this with theories from other disciplines, such as sociology and social anthropology. She also uses environmental climate studies to get details and ideas about life during this time period.

Regarding the information resources that she uses for her research, she explains that stone tools on the settlement sites are her principal materials and that her basic data would be the artifacts and environmental data. She uses artifacts, other objects, monuments, and settlement sites, along with published materials, e.g., articles, books, and archaeological reports on earlier excavated and investigated monuments. She also uses drawings or photos of artifacts in published materials to make comparisons, but not as basic information. Geographical and positional data, e.g., sea level, artifact dating, etc., are the other kinds of information that she uses.

The information systems that she uses in her research include Oria, Google, Google Scholar, Academia.edu (she explains that is because most archaeologists are in Academia.edu.), Musit (the national museum database for archaeological monuments), and Askeladden (Directorate for Cultural Heritage). She uses Musit and Askeladden (archive databases) only as reference sources to find fact-based information, e.g., positions or dating. She satisfies most of her information needs through publications and even has her own database, where she keeps all the information related to her own archaeological findings about studied artifacts.

This participant mainly employs Google Scholar or Google to search for topics. Whenever she wants to search for an author or specific title, she uses Oria. She searches for books when they are cited in other articles to check whether the topics discussed in them are explored more broadly. As for using information systems, she believes that she has a basic understanding of the logic behind the search process and is aware of systems’ limitations. When this participant wants to search archive databases such as Musit for archaeological monuments, she enters a street address, farm (property) name (names can change over time, so it is not easy to use them), farm (property)

number, or else she browse the map. She explains that the map shows all the archaeological monuments, so you can examine them more thoroughly to determine monument type, texture (description and history), descriptions, periods of origin (e.g., the Bronze Age), and curved pictures, if any are available. However, she believes that she can get more detailed information from published articles and reports than from databases such as Musit and Askeladden, which focus on artifacts and monuments.

As for search-results lists, she believes that the further you go in the results, the less relevant listings become. “I normally won’t go too far back because it’s . . . it will be less and less relevant. . .”

To select relevant results, this participant considers the following criteria: titles that contain the search terms; aboutness, i.e., “. . . if there’s any one that has discussed this in particular”; type of resource; type of subject area (discipline); well-known publications, i.e., “if it’s published in one of the large publications”; publication year; abstract; introduction; and where it has been published to see whether there is open access. Sometimes, the results are not directly relevant, but include interesting discussions that she should be aware of on the topic. She said that having online access to a large volume of information – once merely a dream – is becoming a reality. She likes systems with different search possibilities (searchable entities), e.g., the professional version of Musit. She finds it very useful, both for research and for managing linkages to related databases, e.g., linking the two Musit databases, which are concerned with artifacts, with Askeladden, which is concerned with monuments.

#### 4.3.2 Query 3.1: *Evolution Social System*

**Information need:** *“if there’s any one that has discussed this in particular”*

In this instance, the participant, who is working on the theoretical part of her thesis, is going to borrow some theories from disciplines other than archaeology. In this case, she chooses Google Scholar for her search and submits a query on the phrase “Evolution Social System”. She is seeking information and discussions on this topic, i.e., “whether there’s anyone who has discussed this in particular”. She scans the results for titles, resource types, publication years, publications, subject domains, and abstracts to find relevant results. She skips the ones that are completely out of her related subject domains. The participant finally selects a title, “Cultural evolution: social rule systems, selection, and human agency”, a research article by Tom R. Burns and Thomas Dietz. She clicks on the article headline on the search-results list, which takes her to the website of the journal that published the article. She does not know anything about co-author Burns. Whenever this participant confronts a new author, according to her, she likes to find out everything she can about that person.

In this instance, the participant actually is looking for information resources about the searched concept.

**Information need: *author's other publications***

Following her previous search, while she is still at the journal's website, the participant clicks on the "Tom R. Burns" author link to look for any other publications by this author. She would prefer to do this search on Google or Oria, where she believes she would find more results on his publications.

In this instance, the participant is looking for information resources.

**4.3.3 Query 3.2: *Tom R. Burns*****Information need: "*what is the background of this guy?*"**

Following her need for more information on this article's co-author, the participant performs a new search for "Tom R. Burns" on Google. On the results list, she finds a Wikipedia page on Burns and clicks on the link to get more data on him, including biography, professional affiliations, etc. She then extends her search to the university with which Burns is affiliated to find more information about him. Through the university's website, she hopes to find the author's biography, publications, co-workers, co-authors, and other background and research data.

when I start to look into this, the works of this guy, I also of course I checked Wikipedia about what is said about Tom Burns. Yeah so I get the biographical [information] on that.

so even if it's Wikipedia I get an idea. [...] I can read something okay I find out that he is an American Swedish sociologist at the University of Uppsala. So I will go to the University of Uppsala to find examples. So I will go like this, and Tom Burns . . . yeah and here you can find [him]. So, he's working here still, some sort, you can look at the biography and the publications of course. So, that is it. It could be like an entrance sometimes yeah. Because as you noted, any research on regular Google I wouldn't get University of Uppsala. So, then you have to sort of try to find out more of this guy, okay, who it is, who he's been working together with, what kind of work has he done and so on yeah.

This is because I didn't know about this guy before, so I have to make a thorough investigation and what is the background of this guy. So, here it is. Of course you can get all sorts of titles here he's published all the way back to the 80's.

In this instance, the participant is looking for all information about Burns as an entity.

### 4.3.4 Query 3.3: *Tom R. Burns*

**Information need:** “*what kind of discussion has [there] been on this?*”

In addition to information about Burns and his other publications, the participant wants to know about any discussions around the retrieved article, including critiques against it, if any exist. Thus, she is interested in reviews of the article.

[...] I will see if this is, what kind of discussion has been on this, what is the critique for example against this, yeah, when I read this article. So I could also find some critique if I search on this name for example

With the intention of finding discussions and critiques on the selected article, the participant conducts a new search on Burns using Oria. She gets results on two “Tom R. Burns” entities. One works in sociology and the other in psychology, which is not in her field. Thus, she refines her query on Burns. (She possibly could have narrowed the results by resource type to “reviews” in the search system, but she did not.)

Browsing the results list, the participant gets excited after finding an article that is a kind of review, titled “Other reviews: The shaping of social organization”. (It is worth noting that the article’s resource type identifies it as an “article”, not a “review”, which means that she would not get access to this review if she already narrowed the results by resource type to reviews. Maybe this was why she did not limit her search to “reviews” as resource type.) In addition to this review article, she found other articles that are not considered relevant to this search, but are relevant to her other information need regarding her conference paper, including “Actor-system-dynamics theory,” a topic that interests her, and “Technology, Complexity, and Risk”, which she believes is relevant and quite new. She selects another title because its author co-authored another article with Burns. The participant believes that by delving further into the results pages, she will encounter irrelevant results on the other “Tom R. Burns”. She said it would be interesting just to know whether one specific theory is still being used and discussed, although she has no plans to read articles about it.

In this instance, the participant is looking for information resources that are about the retrieved article.

## 4.4 Story 4: User is interested in information about translating of books

### 4.4.1 Introduction

This participant is a female in her 30s and a PhD candidate in translation studies. Her project is about the sociology of translation and agent studies (i.e., the people who directly/indirectly impact

translated texts). She explains that she has background knowledge in translation studies provided through her bachelor's and master's degrees in this field, as well as her field work as a translator for a few years. She uses different kinds of information systems and databases to seek required information for her project, including Oria, Google, Academia, John Benjamin's Translation Studies Bibliography (a bibliography database for translators), and ATEKST (Retriever), a media-archive database that contains all newspaper and other news media articles in Norway. She is quite satisfied with the results yielded and does not usually filter results by provided categories or options.

I think the results page I think it's good, sometimes there are too many options to filter the results and I don't always use them because I don't have the time to, I don't take the time to do it because I'm lazy.

I don't want to rearrange the results because just to get through the irrelevant ones or no because it's, I can see it doesn't take me any time. I can see it straight away, I see that this is not the same book, or even if I just click into this page just to make sure it's not relevant, it doesn't, okay so this wasn't relevant so I just close it. The ones that are relevant like this one okay, I save them, save as and then I put them in my computer.

#### **4.4.2 Query 4.1: Actor Network Theory**

**Information need: *the English version translation of a the book that she already has in Spanish book***

The user is looking for specific book in an English-language version of a book on that is about a particular theory called actor-network theory. She already has the same book in Spanish, but she is writing a book and wants to use English quotes from the book in her book. Therefore, to find the book, she performs a query search on "actor-network theory" on Google that contains part of the book title, *Reassembling the Social: an Introduction to Actor-Network Theory*. However, it seems that this is the principal part of the title and topic. When she browses the results list, many contain this particular theory in their titles, but they are not relevant because they are in other disciplines or have the wrong date.

so that's applied to all kinds of different things and I'm not interested in researching education, I'm not interested in information systems, that's when I know that is the title I'm looking for. This is an introduction, "Reassembly the social, an introduction to Actor Network Theory", but I actually have this book already in Spanish, but I want to revisiting it to quote it in my work because I don't want to quote it in Spanish.

In this instance, the participant is looking for a particular information resource.

#### 4.4.3 Query 4.2: “Jennifer Egan Bakke”

**Information need:** *What did reviews say about the translator?*

In this example, the participant is working on a translated book titled *Bølle på døra*, written by Jennifer Egan. The participant explains that she wants to know what has been said about the book’s translator, Kyrre Haugen Bakke, and his translation, i.e., the user is looking for some reviews and critiques about the quality of this particular book’s translation. Therefore, the participant performs a search task on ATEKST Retriever, a news archive in Norway. She queries the phrase “Jennifer Egan Bakke”, which contains the author and translator’s names. In the results list, the participant finds a review on Bakke’s translation work on *Bølle på døra*. She also finds another text in which Bakke responds to a critical review of the book by Bernhard Ellefsen. The participant seems excited about finding these items.

... so this is Kyrre Haugen Bakke. He’s a translator and he’s referring to this article which I can also find and retrieve it, okay, where he’s, there is a reviewer who is criticizing his work and so this is the translator’s reply to what the critic wrote here. So what I did was I read, here I’ll just get one page okay, but I read all the reviews about this book by Jennifer Egan.

In this instance, the participant is looking for information resources.

#### 4.4.4 Query 4.3: “David Vann Jord”

**Information need:** *“What did reviews say about the translator?”*

In this example, the user performs a search task on ATEKST retriever, which is a Norwegian news-archive database. Her query, “David Vann Jord”, combines the author’s name, David Vann, and the title of his translated book, Jord. The participant said she hopes to find reviews and critiques about this translated book in the newspaper. She is interested to see what people have said about the translator and the translated book.

so that’s the name of the translator (Hilde Stubhaug), the name of the author (David Vann), the name of the book (Jord), the name of the newspaper and so then I would keep this one and I would go to the next one. So that’s written by “Katrina Krøger”, so the same book, the same title, same author, same translator, same publisher but it’s a new review. So then and what I’m especially looking for here is what they say about the translator okay, but I would read the whole review in each case.

The user finds another article, “Mørkt som i graven” (In English: “Dark as the grave”). This article is a review that criticizes Jord – another relevant result for the user. “I would read this one because this is a review, and it mentions the translator.”

In this instance, the participant is looking for the context around a translated book.

## 4.5 Story 5: User is interested in information about a particular travel writer

### 4.5.1 Introduction

The participant is a female in her 30s and a literature PhD candidate. She is halfway done with her PhD (nearing the end of her second year) and has been working on comparative literature studies for the past five or six years. Her PhD project also is on comparative literature and deals with Canadian travel writing. This participant believes that she is working in a multidisciplinary context because besides her principal discipline of comparative literature, she is using publications from other disciplines – e.g., anthropology, ethnography, cultural studies, literary sociology, and political science – for perspective and background.

She mainly uses books, book chapters, articles, and archive materials. The different information systems that this participant mostly uses include Oria, JSTOR, Modern Language Association (MLA), and online archives from university libraries (mostly Canadian university libraries, e.g., Library and Archives Canada [LAC] and McMaster University Archive). She usually begins her searches in Oria because of its accessibility at the university. She understands what she can look for in each database. “It (McMaster University Archive) is a database that is used a lot in humanities, especially English literature”, she says.

Being aware of the limitations and incompleteness of indexing, the participant uses advanced searches that include truncation, word combinations, and synonyms in her queries to both broaden and narrow her searches to find everything she needs. Depending on the search, the participant narrows her query by language, publishing date, and material type (e.g., books). However, among the categories within the results, she finds subject categories to be useless and does not use the suggestions offered for new searches. Generally, this participant prefers information systems that provide the most relevant results in the first two to three pages. She trusts these systems and the way they organize and present the information. “You have to trust libraries to present (data) in the best way, and I do. . . (though) it would be better to have more subject words”. For this participant, having search terms marked or in bold type does not help with finding relevant items. She only cares about publication year if it concerns articles, or if a book has different editions, to help her choose the newest edition. Otherwise, the title and subject are more important to her in selecting relevant results.

I know that the early 1990s is also important in travel writing field, but if the title and the subject is interesting enough, I don't care about the time. But of course if there are different edition too, I try to find the newest one, in case there is an extra foreword or additional chapter or all those things, that is also important.

This user prefers information systems that provide the most information on results lists and the shortest paths to the information she needs (i.e., not requiring several clicks to find it).

So, that would also a kind of wishes to have everything in one search and to get everything listed to see where everything is at once but you don't have to click to... I think that is what I can think of, basically in a way more information at once, on one site not so much clicking.

Therefore, she prefers an integrated search system in which she can get several different kinds of information related to search terms through one search, in one search system, instead of needing to use several databases and search systems for different types of materials.

It is a display of the different kinds of information that is relate to the person or the subject or place that you are interested in rather than having to go to start the other way around, to start with the text, start with image, start with the [...].

The user likens making the same query in different databases to opening several different doors in the hope of finding the needed information and going back and forth through these doors, instead of having them all opened together.

Also a way of visualizing it. Because that is important too. Now I feel like I am going into like the small door and when nothing is here, I close the door and go back and the next door, rather than have it all laid out for me.

Regarding the representation of information in an integrated system, the user imagines it as a tree that provides her with an overview of all related information from all providers. Depending on what she searches for, she prefers that the information be organized based on type of material, then date and time if she is searching for a person. When using a keyword, she prefers that the information be organized based on discipline first.

It is organized of course but you have to go to all of these different levels. You don't get it all at once. I am imagining some kind of tree almost, ok they all have the photographs, they all have the text, they have official documents, they all have the books.

On the results page, she prefers to have more information and access to content from all information objects through subject words, allowing her to see cover-page images and abstracts or other summaries of each result on the results page. Since titles are not always informative enough, she would prefer to have discipline information (field of study) as well, along with a short summary, accompanying the title and subject:

“Of course, [with the] Journal of Folklore Research, we know what discipline it is, but [with] journal articles or book titles, they are not really clear. So, it is not

easy to guess what research field that is. And if they are not called ‘studies’ and ‘travel writing,’ but something else, yes, this summary is also just a small kind of disciplinary label and, of course, more if there is a need for that.”

This participant prefers archival documents in paper formats. She says having physical documents makes it easier for her to decide which documents are relevant. This is due to not all archival documents being digitized. Moreover, the descriptive information in archival documents is insufficient for her needs as far as providing document content.

when I look through these online things in a way I don’t feel that I have control. Because, of course I get the impression but I do not have the physical things to see and to decide for myself[. . .] And I think that was easier when I have all that material I can sit there, go back and read it. Because the description is not enough [in archival databases].

because today at least you will always be able to access maybe a part of a book, a part of article online, Google books, Amazon or that at least to get a glimpse of it. But for these materials [archival materials]there is no way to do that. I think it is much more mysterious to work with these kinds of archival materials.

#### **4.5.2 Query 5.1: *Travel writing and Canad\* and North***

**Information need: “a book about Canadian travel writing in the North in general”**

In this instance, the participant wants to find “a book about Canadian travel writing in the North in general”, but she doesn’t believe such a book exists.

My favorite ideal could be a book about Canadian travel writing in the North in general, but that does not exist.

She starts a search on Oria and queries a combination of keywords: “travel and Canad\* and North”. She narrows the results by choosing “English” as the language and “books” as the material type. She also tries to set time parameters, but it does not work (maybe because of indexing-data or search-system shortcomings). She gets several hits from other disciplines that are not relevant to her information needs and refines her query by adding one more keyword to make it more specific: “travel writing and Canad\* and North”. This is a type of topical search, in which place name is used as a topic.

Among the 48 results, browsing the first 10 results, two catch her attention, one of which is an article titled “Riding/writing across borders: North American Travelogues and Fiction”, written by Tom F. Wright and published in *Studies in Travel Writing*, a relevant and known publication

for this user. The other is a book titled “Through Femenist Eyes: Essays on Canadian Women’s History”, which a colleague recommended. In this book, only one chapter seems relevant to her needs and is about travel, the North, and the Inuit, all of her topic interests. She notes that she would not have found this book (chapter) if her colleague had not recommended it to her because the indexed subjects for this book (“Canada, Western–History”; “Women–Canada, Western–History”; and “Women pioneers–Canada, Western–History”) were not informative enough for the participant.

Because one of the chapters is about the travel and north and Inuits (Eskimo). That is interesting why it has not been listed as subject. Maybe if I had not been recommended it, maybe I would skip it. Just because of the insufficient subject words.

One of the other results is a review article about a book titled “Riding/ writing across borders: North American travelogues and fiction” that seems to interest the user. She also gravitates toward the reviews because she says reviews usually are short and give her an overview of the entire book, which she is not sure she will read. She finds reviews to be very useful for books that are not available online and compares reviews to academic articles’ abstracts.

I find that (useful) because usually it is only a couple of pages that gives you an impression of the book and it is really useful for those books that are not available through the university library. I can at least try to picture ok if it is something that I should order online or ...and with the articles I read the summary and if it is interesting, I read the whole article.

In this instance, the participant is looking for information resources about the searched topic.

### 4.5.3 Query 5.2: *Richard Harrington*

#### **Information need: *copyright holder of photo***

As part of her project, the participant is writing a book in which she wants to reproduce a particular photo that a travel writer shot during his travels to the North. This travel writer took many photos with groups of indigenous people while traveling to different places. In this instance, the participant has an image of the photo, but needs to know who owns the legal rights so she can ask for permission to use it in her book.

These are the photographs by him and the reason that I had to look it up because I needed to find, it was actually the practical thing, but I needed to find the copy right holder of the photograph.

... Because even though I am just going to reproduce photographs from the book, I need to ask permission for those who actually own the photographs which is the LAC [Library and Archive of Canada].

So, for this information need, which is a kind of fact-finding search, the participant performs a Library and Archives Canada image search, which employs a very simple search using keywords and material type. The user inputs a query on the photographer's name, "Richard Harrington", and it retrieves over 1,000 images. She then tries to narrow the search and adds both the site (a copper mine) and the name of the specific group of people (Inuit) in the picture. With such a specific query, very few results show up, which could be a good thing, but none is the photo that the participant seeks. She is aware that this is because the archive has not finished indexing the images, resulting in insufficient bibliographical information. Thus, she returns to her previous general query, which was "Richard Harrington", and browses through the hundreds of results manually to find the image, then the copyright holder, then the holder's contact data.

The participant in this instance is looking for some fact-based information about a picture.

#### **Information need: names of the people in the photograph**

This participant already knows that this archive also has started the project of identifying the names of people in photos and wants to find information about the people in the photo (if such information is available) to add to a book footnote.

I think they have a project which they called project naming which is about identifying all the specially the Inuit people who they are not identified. So, they use elders from the Inuit communities to get them to identified the different by name.

[... ]and I know that they are doing this for the Richard Harrington's photographs. But they are not done yet, I think it is difficult. Since this is from the 1950s and they need to find the relatives.

[... ] I actually think that if they have done it, I would probably have a footnote with the names just to kind of not to reproduced the stereotypes in a way, but a kind of make use of this work. That is just a detail but I found it interesting.

Like the previous instance, the participant seeks some fact-based information about a picture.

#### **4.5.4 Query 5.3: *Richard Harrington***

##### **Information need: *all related textual information and documents on his journey***

In this instance, the participant wants to find all related textual information and documents on Harrington's journey to the North. With this information need, she decides to use Library and

Archives Canada's Archives Search system and inputs a query on "Richard Harrington". Since the participant is interested in only texts this time, she narrows the search to "textual material" as the material type. She believes that in an archives database, she could find many different kinds of documents such as the photos that Harrington took in his journey, newspapers and other sources.

On the results list, "Richard Harrington Fonds (textual record, graphic material)" is the first item. She clicks on it, and in the description, it says that it contains only 5 cm of text records, which is not what she wanted. Thus, the fonds are mainly photos. The fonds' hierarchical structure is a very long list of 41 pages without any specific order, so it would not be easy to browse. She admits that it was good that the fonds content was categorized based on material type.

Q: So, it would be nice if at least it is categorized based on the text, images, . . . ?

Resp.: Exactly, like you have folders on your computer that you have folders of photos, text then the,

Q: Then the list of each of them?

Resp.: Yes, just to get the more of an overview since there is so much material.

In the end, she went back to the results list to look for texts. The fifth item on the results list interests her, as it is the title of one of the texts about which she is writing. The document is titled "Harrington, Richard (The Face of the Arctic)". When she clicks on it, she finds that it is part of a file named "Central registry files (textual record)", which has a very long hierarchical structure (more than the system could show). As the participant states, based on her previous experience with the archive, this extensive file contains governmental documents concerning Harrington's journey, such as government permits, transportation, police services, and many other items in which this user is interested. The participant refers to these as "department records", and she aims to find Harrington's files. No access to the textual materials' digital files is available in this database.

The hierarchical structure helps her see what the item is and what it contains, such as sub-sub-series, then she can find the reference number and query the archive, which will find it for her.

[. . .] I have to send in this reference no. and everything. So they of course would find it for me. Then, I didn't have to go through a lot of material.

The Face of the Arctic is a travel book in which Harrington writes about his journey and is full of photos alongside the text. However, what she found here (in the archive database) is not the book itself, but some documents about his journey. This descriptive information is not enough for her needs. To see the book, she must do another library-database search.

[...] But the thing is that if I wanted to find the book itself, I had to go to the library section and search there. Because here it would only be kind of documents and it might also be editorial documents, you know. This is not here but since this is the national archive, they would have everything.

When the participant is asked whether she is interested in the book itself, she explains that she is interested in “Richard Harrington’s journey” itself and all existing documents tied to it:

Q: But here you are interested in the book itself?

Resp.: Yes, the kind of the journey itself. About the journeys that he based his book on. And then I was interested in kind of way he funded it and how he, what the government [thought]of it. what were their relation and correspondence and everything? That I found there which was really interesting. I think it was also interesting in my theses to kind of draw a context again.

Q: Not only the journey but?

Resp.: yes, everything around; not only show the single person but you need permission, you need money, and because also the cold war, there is also a sense of censorship. Because I know that Harrington, he wanted to go to radio station where both American and Canadian forces were, but he was not allowed. Because they were afraid of spies and. . . . So, that are things that I have learned from these files.

In this exploratory search on Harrington’s journey, in this instance, the participant is looking for the full context.

#### **4.5.5 Query 5.4: “Richard Harrington” AND “face of the arctic”**

##### **Information need: all literature on his travel book**

In this instance, the participant wants to know what people have said about this person and his travel book, and how this book has been used as a source in other domains. Thus, the participant performs an advanced search in JSTOR and query on “Richard Harrington” AND “Face of the Arctic”.

The results list shows two items: One is a research article published in the Journal of Central Anthropology and is not very relevant to the participant’s information needs because it is not about Harrington’s book. This article only cites the book within the text. However, it is interesting for her to see how travel books are being used and if they have been used as novels or as scientific resources.

But this is not about his book. But what is interesting to me about this is that I know that he is referred to the anthropology studies not too long after the war. That is

also kind of a way that I use Jstor just to find out how the travel books are uses as sources, not as fiction or personal narratives but as a knowledge in a way.

The other result is a review article published in Geographical Journal, and she finds it to be relevant because it is about the book.

In this instance, the participant is looking for information resources.

## 4.6 Story 6: User wants to learn about value chain

### 4.6.1 Introduction

The participant is a man in his 30s who started his PhD project on medicinal plants about a year and a half ago. He has been working in the field of molecular evolution and bioinformatics for about 10 years. The information systems that he uses include Google, Wikipedia (for examples and to find a definition of a concept), Google Scholar, and Web of Science. Regarding informative material and resources, he uses mainly journal articles. He does not use books that much, but does use manuscripts. “So, I’m looking for manuscripts before the 11th century to track some data that traces the trade of this plant during this period or before this period.” As for results presentation, he believes that “a list is not really intuitive” because you need to go through all the titles.

### 4.6.2 Query 6.1: *value chain*

#### Information need: What is value chain?

The user is working on a particular concept known as value chain in his research. First, he wants to know what a value chain is, its aspects, and the concept behind it. Therefore, he submits a query on “value chain” in Google. The Wikipedia page on “value chain” contains his desired results, but the participant still wants an abstract graphical representation of the concept.

This participant seeks information on a particular concept, “value chain,” as an entity.

### 4.6.3 Query 6.2: *value chain*

#### Information need: graphical abstract overview

After finding some information about the concept of “value chain” on its Wikipedia page, the participant also wants to find a flowchart on this concept to give him an overview of different aspects on it. Therefore, he performs a new search and inputs a “value chain” query on Google’s image search.

so I do that first to see what does it mean . . . I could go on Wikipedia to understand what the definition, I mean it’s always good to find that, and sometimes I go to

images because sometimes you have a work-flow with that and I think it's, I think it is helping to understand . . . the conceptual, the concept behind the value chain. I think it's very interesting because now you have the availability of an abstract or a view of a different aspect of value chain from, it could be from a company and with contributing different aspects of the value chain. . .

He clicks on a couple of retrieved images, examines them, and visits the original web pages where the images originated. He gets an overview of the concept, but is still not completely satisfied and needs more information.

In this instance, the participant actually is looking for information resources, or rather, a particular component of an information resource.

#### **4.6.4 Query 6.3: *value chain***

**Information need:** “[...] *synthesis of the value chain; so, a figure*[...]”

Following up on his previous search to find a graphical representation of “value chain” and its different aspects, the participant performs another search on this concept in another database. This time, he uses Google Scholar to search for the concept. He still is seeking a flowchart or another graphical abstract overview of the concept that he hopes to find in research articles.

I am looking for something that is synthesis of the value chain; so, a figure in [an] article that would be an abstract of this different information.

He finally finds an article that includes a flowchart on the concept of value chain and its related workflow.

The participant in this instance is looking for some part of an information resource.

### **4.7 Story 7: User wants to know about gender and rights to the city**

#### **4.7.1 Introduction**

This participant is a female in her 30s and a social-sciences PhD candidate with a multidisciplinary background in environmental science and social sciences. Her project is about cultural geography, urban geography, psycho geography, graffiti, and street art.

The information resources that she uses are mostly books, academic articles, journal articles, and her own photos. She usually searches for information in Google Scholar (“to look up specific articles”), Google (“book, . . .”), and Academia. “So, mostly, when I use Google Scholar, I’m almost always looking for journal articles. If I want a book or something, I will just use Google.”

She believes that since she is doing interdisciplinary research, she must look for information in many different databases to avoid missing certain results. She explains that

my work is so interdisciplinary that it was very difficult, I would have to look in so many different databases [...] and my work is a little bit of this and a little bit of that. I would often be missing certain results.

, sometimes you don't want it to be too relevant or too restrictive, then you don't get, if you want to do interdisciplinary research, then you want to have access to sort of outside of your field. That's a problem I used to experience a lot with the databases, they were so restricted by discipline that if I was looking in a geographic one, I wouldn't have access to anything in art history or the art journals, which are very relevant for my work.

She explains that when it comes to new topics that she wants to research, she asks colleagues and others for literature suggestions to help give her a place to start.

So that's often what I'll do, actually, is I'll talk to colleagues or people and ask for suggestions of literature, especially if it's something that I'm not that familiar with, the gender issues is something new to me and sometimes it's difficult to know where to start, which authors to look at. So often I will look for recommendations from other colleagues or friends.

When she gets a results list, she usually browses the first three to five, or sometimes 10, pages for relevant results. However, she believes that "they put the most relevant, or the most-cited anyway, first". To find relevant results, she looks at journal titles because certain journals are more relevant to her field than others. Moreover, she has a list of favorite authors who share her research interests.

She is quite happy with the simplicity of Google Scholar's results list and thinks too many categories would be problematic. She prefers a simple results list that shows which results offer full text, etc.

#### **4.7.2 Query 7.1: *Gender and Power Connell***

**Information need: *a Connell article on gender and power***

In this example, the user is looking for a particular journal article. She knows the author and part of the article's title; therefore, the user performs a query on Google Scholar with the phrase "gender and power Connell". Her query comprises the article's incomplete title and the author's surname.

I was looking for an article by Connell on gender and power. It was a recommendation from somebody. So often it's kind of specific articles and an author that I know they've written and can't remember.

She discovers that the full title actually is that of a book, *Gender and Power: Society, the Person and Sexual Politics*, by R.W. Connell. In this case, the user was looking for a specific known article by Connell. However, having found this particular book, she is still interested in examining Connell's other works on this topic, as she knows that Connell has written extensively on this topic.

He's written a lot on this topic so I was looking not for a particular one, just so see different things he had written.

In this instance, the participant is looking for a particular information resource.

### **4.7.3 Query 7.2: *women right to the city***

#### **Information need: *specific article***

In this instance, the participant is looking for a particular journal article about the topic of women's right to the city. She inputs the phrase "women right to the city" into Google Scholar and browses the results list. A couple of titles capture her attention, one of which is an article by a well-known author, "Mitchell", whose works she is interested in, and the other is an article that she read about in another paper. Then, she finds an article titled "The Right to the Gendered City: Different Formations of Belonging in Everyday Life", by T. Fenster. This is exactly what she was looking for. To explore this topic further, she consults this article's references, which she believes is "the best place to find the resources, especially if it's a really, really relevant article, and well-written". She has found that Fenster has written extensively on this topic and is interested in reading all her works, a list of which she seeks at Academia, including published journal articles, unofficially published articles, works in progress, conference papers, etc.

The participant in this instance is looking for information resources in particular.

## **4.8 Story 8: User is interested in archaeology**

### **4.8.1 Introduction**

This participant is a woman in her 30s who just finished her PhD project in archaeology, about houses during the Viking Age. The type of information resources she usually uses are unpublished excavation reports, plan drawings, journal articles, and newspaper articles. She usually finds these information resources in archives, regional museums, topographical archives, regional archaeological journals, archaeological blogs, Google, Google Scholar, and library

databases. Other disciplines that this participant uses for information include social anthropology and sociology, which she uses for theoretical frameworks, and natural and Earth sciences. In research articles, this participant says the most important parts are theoretical framework, method, and principal argument, which usually can be found in the introduction or conclusion and the references sections of articles, with the number of articles cited indicating how successful an article is.

Because of her field and the kind of research she is doing, the information type she is looking for is very specific to the field, e.g., “What farm is it, what’s the farm number, which county, which municipality?” She also explains what kind of information her ideal database would contain:

[. . .]so how many hearths or fireplaces are there, are many rooms are there? Can we estimate how many people lived in this house? Can we estimate how they used different rooms, how many doors are there? What does that tell us about access and exclusion, if there were?

As an archaeologist, her principal problem is finding primary materials because no principal database for them exists. She states that:

So as an archaeologist the main problem is getting the primary, the primary material which is not so, which there is no central index or database of and you need to know people and you need to go through lots and lots of catalogs to be able to sort of find things that can take years you know if it’s a big set of material.

#### 4.8.2 Query 8.1: *materiality archaeology*

##### **Information need: *How it has been used in archaeology?***

In this instance, the participant, working on the theoretical framework of her project, wants to know about a particular concept – in this case, how materiality has been used in the field of archaeology. The participant performs a search task on Google Scholar and enters a query on “materiality archaeology”, which is a combination of the concept and her discipline (research field).

if I wanted to use different theoretical frameworks. I could definitely just do a Google Scholar search. It’s a big theoretical movement now about materiality . . . so if I wanted to see how that’s been used in archaeology this is all of the social sciences in the humanities, this is sort of a big wave that’s coming. If I want to see how that’s been done in archaeology I would just do something like this, materiality, archaeology, yeah, and then see, I got 21,000 results.

On the results list, many results have “materiality” in their titles, but in different geographical contexts. One is Scandinavian, which is relevant for the participant, a book titled *Iron Age Myth and Materiality: an archaeology of Scandinavia AD 400-1000* by Lotte Hedeager.

of course not everything might be relevant but you see that most of these do have materiality in the title and they're just from very different places and geographic context. This is Ghana materiality and this is a whole sort of collective book about materiality. Middle East, Eastern Mediterranean, this is Iron Age, this is in Scandinavia.

Then she finds an article written by a known author. The user is interested to see how many times the article has been cited because she wants to know how popular it has been. Moreover, she wants to read other publications by this author on this topic, which she easily can find by clicking on the name of the author in Google Scholar as she explain:

[...] I want to see what is she writing about this topic, I know that she's doing good things, I want to know what she's written about materiality and the body, and then I would click her name and find her publications. That's the easiest way to do it.

The participant in this instance is looking for information resources in particular.

## **4.9 Story 9: User is looking for information about development of children's libraries**

### **4.9.1 Introduction**

This participant is a female in her 50s who is a final-year library and information-science PhD candidate. Her project concerns development of Norwegian children's libraries, and she uses mostly books and scholarly journal articles, archival documents, newspaper articles, annual reports, architectural plans, and pictures and photos as references. The information systems and databases that she usually uses include Oria and EBSCO.

In search results, this participant wants to see some meta-information about retrieved documents or articles, which she lists as “author, (a very descriptive) title, abstract, references, year, publication (publisher), and, of course, the keywords”. She also likes to note the frequency of search terms in documents to get an idea of how relevant the document is. She also wants to know the number of pictures or photos in the document, if any.

I would like to see how frequent the name or keyword appears in this document. It will give you the amount of relevance. I would also like to know if there are pictures or photos related to that. [...] or it can just tell me there are 7 pictures in this article.

#### 4.9.2 Query 9.1: *sociology AND childhood AND Giddens*

**Information need: "How [do] they use his theories?"**

In this instance, the participant already knows about a sociologist named Giddens, whose theories she uses in her work. Actually, the theoretical part of her work is based mainly on this sociologist's theories. She wants to know how other scholars have used Giddens' theories in their works. In fact, she wants to prove her methodology's validity. Here, she explains what she is looking for:

Q: So, you have searched for Giddens' works here?

Resp.: He has not written so much about childhood but I am looking for the other sociologists or LIS researchers who use his thinking in their studies of childhood[. . .]. I wanted to know if anybody has used his insight. (p. 4)

[. . .]It would be more interesting to see how they use his theories in a way that is more like the way I use him. So, I would like to have some confirmation that the way I use his concept is ok. So I am kind of looking for backing up.

To meet her information needs, the participant inputs a query on "sociology AND childhood AND Giddens" in the EBSCO database. Her query contains some concepts, plus the sociologist's surname.

Here, the participant's intention in searching for the sociologist's name is not to find any information about the sociologist himself, nor his theories, books, or articles concerning those concepts. She is not even interested in works that others have written about this sociologist and his theories. Thus, it is not easy to discern the user's objectives merely through her query. The user's actual information need in this case concerns how this sociologist's theories have been used in other research about a particular topic that other researchers have examined in other specific domains. In fact, the use of Giddens' theories in particular topic areas in others' work is her principal interest.

The participant in this instance is looking for information resources in particular.

**Information need: "In what way are they [children] agents?"**

Through this search, this participant also expects to find "the outcome of the Giddens theories". She wants to see "how they work with children as agents or as library users [. . .] if they are demanding services or just humble and quite. Do they stand up for their rights in a library? In what way are they agents?"

In this case, the participant is looking for information resources.

Except for a couple of searches that this participant performs, she mentions some of her information needs that seem to be very difficult or impossible to search for. The following information needs are of this kind.

### **4.9.3 Information needs sought manually**

#### **Information need: “is it made just for sitting down or is it made for running around?”**

As part of her research, this participant is looking for pictures of libraries, including their interiors, to get some ideas about children and libraries in the early 20th century. However, she has tried to find such pictures manually because she found it difficult to search for them online.

I was looking for pictures and information about how children libraries were; how their interiors were changing in the first part of 20th century. [. . . To see] how they [library buildings] changed throughout time. So, I would like to have both photos and I would like to have plan drawings from libraries which it turns out to be very hard to find. Because there is a connection between how we think about the children and how we plan and equip a library department. Is it made just for sitting down or is it made for running around and being comfortable? Is it just books, shelves and tables? Or is it play ground? it has something to do with the view of children and the view of what a library should be. So visual representations are very interesting for me.

Here, the participant is looking for some ideas and context.

#### **Information need: “whether cultural clashes were documented or discussed”**

The participant in this instance is interested in “how Norwegian children’s libraries have adopted the American library traditions”. She believes that Norwegian society and U.S. society are very different in ideas and politics. She is looking for discussions around the probable war of ideas.

Somehow this American ideas fitted in nicely in spite of clashes that was evident. I wanted to find out if those clashes were documented or talked about or discussed in journals or in newspaper articles inside the profession though.

However, she found it difficult to find such information online and had to do it manually. “That was impossible to search for. I cannot search for it. I just pick a bunch of newspapers or potential journals and read through them and make notes.” Even if online searches were an option for her information needs, she still prefers to look for such discussions manually – in this case, discussions around adopting U.S. traditions for developing Norwegian children’s libraries – because she wants to determine the context around the discussions as well. She believes that even

if she could search online, the retrieved results would yield a couple of articles from different journals that stand alone, lacking contextual information. “If I could search for that, it would have saved me some time, but on the other hand, I would have missed out on the context going on in these discussions.” She further explains that by reading through different journals, even by looking at the ads in each journal, she can get the desired context around the discussion.

Resp.: I would need the context for the discussion. By going through all the years of this specific library journal, I could see the context for the discussion. So, it gave me some more than just those articles.

Q: What if the system can give you the other articles that have been published in the same issue of the journal?

Resp.: It is hard to say. When you go to historical documents, it is also. . . you get in touch with the history by the layout, the feeling of the history, I mean. This is not very scientific. [Showing me a physical library journal she continues:] This is a very “modern” journal from 1933, you can smell the history. . . Even I could see the advertisement which tells you a lot about the time and the way they think about [ . . . ]

Q: What are these advertisements about?

Resp.: All kinds of things: coats, military uniform, painting, bank, some equipment for stopping the fire, . . . . It is just the picture of the time. Not that I have been studying all those advertisements very much but it gives you a sense of something.

In this instance, the participant is looking for context.

## 4.10 Conclusion

Different instances in this chapter’s stories show that what participants usually query about is not necessarily what they need. Sometimes their actual needs are quite difficult to guess just by examining their queries, which usually are based on people’s names, concepts, or information-object titles. However, much more lies behind their queries that needs to be discovered. Moreover, they sometimes have several connected information needs that they express through different queries, or even similar queries in different information systems. The next chapter will discuss these findings more thoroughly.



## Chapter 5

# **Analysis and Discussion**

During the second and third phases of analysis, participants' extracted information needs and associated contextual information from the collected data have been presented concisely in the holistic model of information needs progression and inclusively through participants' stories about their information needs, presented in the Results chapter. The stories in the previous chapter provide a foundation to better understand the more granular analysis and discussion that this chapter provides.

The results from significant analytical steps will be discussed in this chapter in four parts, starting with a holistic examination of users' information needs, in which I discuss the conceptualization of users' information needs through a holistic approach. A discussion follows on the different types of information needs that have emerged during this study, including fact-finding, resource-finding, entity-finding, and context-finding. Next, I examine general patterns in users' information needs when they search for cultural-heritage information. Emergent patterns have been projected in the ontology CIDOC CRM and its domain-specific extensions. The chapter ends with an introduction of the model of semantic information needs.

## 5.1 Holistic perspective of users' information needs

In the present study, we tried to approach information needs from a holistic perspective, as opposed to conceptualizing information need as one single phenomena. In the holistic perspective, we view users' information needs as a whole, not as a collection of parts. Each part of this whole can be best understood in the context of users' primary information needs and in relation to other interconnected information needs, which we call secondary needs, regarding a particular work task. Using a holistic perspective to conceptualize users' information needs, an information-needs progression model (Fig. 3.1) was developed in the second phase of data analysis. The model was introduced in the methodology chapter as a tool to analyze data collected through contextual interviews in a qualitative study. The results of this phase's analysis have been summarized in Table 5.1, taking an ontological approach. In this table, following the holistic perspective on information needs, participants' information needs have been categorized in three parts: query, i.e., what participants enter into information systems as a compromised expression of their information needs; primary needs, i.e., what participants express as their actual information needs; and secondary needs, i.e., what participants state as other actual needs. It is worth noting that the secondary needs in this context are as important as primary needs. The secondary label is not meant to convey less importance. Secondary needs here refer to information needs that participants express after primary needs, mainly because the participants realize these needs after interacting with information systems, while retrieving information. They also can surface while seeking information under Taylor's information-need levels, eliciting compromised, formalized, conscious, or even visceral needs. The categories introduced here – i.e., query, primary needs, and secondary needs – provide a holistic perspective to analyze users' information needs in qualitative studies using contextual interviews as a data-collection method.

Table 5.1: A compact summary of participant's information needs

Holistic view to participants' information needs					
Query		Primary need		Secondary need	
Entity	Example	Entity	Example	Entity	Example
concept	<i>history of property</i> (Story 1)	information object (about)	<i>secondary resources for theoretical background</i>	<ul style="list-style-type: none"> <li>• Person (as an entity)</li> <li>• information objects (by person)</li> <li>• information objects (about person)</li> <li>• information objects (about information objects)</li> </ul>	<ul style="list-style-type: none"> <li>• <i>facts about him</i></li> <li>• works by him</li> <li>• <i>works about him</i></li> </ul>
concept	<i>intellectual property fashion</i> (Story 1)	information object (about)	<i>secondary resources for theoretical background</i>	information objects (by group)	<i>other works by historians and sociologists</i>
concept	<i>evolution social system</i> (Story 3)	information object (about)	<i>discussion about it</i>	<ul style="list-style-type: none"> <li>• information object (by person)</li> <li>• information objects (about information objects)</li> </ul>	<ul style="list-style-type: none"> <li>• <i>Author's other articles relevant to the topic</i></li> <li>• <i>discussion around the article (reviews and critiques)</i></li> </ul>
concept	<i>Rural employment Guarantee</i> (Story 2)	concept (as an entity)	<ul style="list-style-type: none"> <li>• <i>all about the act</i></li> <li>• <i>its forming and its evaluating</i></li> <li>• <i>its different aspects</i></li> </ul>	context (around a concept)	<ul style="list-style-type: none"> <li>• <i>how people perceive it?</i></li> <li>• <i>how it has worked in practice?</i></li> <li>• <i>how people are interested in the act?</i></li> </ul>
concept	<i>human rights and ecology</i> (Story 2)	<ul style="list-style-type: none"> <li>• information object (about)</li> <li>• person</li> <li>• legal body</li> </ul>	<ul style="list-style-type: none"> <li>• <i>books or articles</i></li> <li>• <i>people who write about it</i></li> </ul>	<ul style="list-style-type: none"> <li>• person</li> <li>• legal body</li> </ul>	<ul style="list-style-type: none"> <li>• <i>to know if they have written 10 articles on the same aspects (how many times they have been writing about the same issue)</i></li> <li>• <i>co-writers</i></li> <li>• <i>who and what is written</i></li> </ul>
concept	<i>Actor Network theory</i> (Story 4)	information object (about)	<i>the English version of a book about</i>		

Continuation of Table 5.1					
Query		Primary need		Secondary need	
Entity	Example	Entity	Example	Entity	Example
concept	<i>value chain</i> (Story 6)	concept (as an entity)	<i>what is value chain?</i>	information object (about)	<i>graphical abstract overview of value chain</i>
concept	<i>value chain</i> (Story 6)	information object (about)	<i>flowchart or workflow</i>		
concept	<i>women right to the city</i> (Story 7)	information object (about)	<i>specific article</i>	information objects (by)	<i>all her works</i>
concept	<i>materiality archaeology</i> (Story 8)	information object (about)	<i>books or articles</i>	information object (by)	<i>other works that the author has done on this topic</i>
concept	<i>travel writing and canad* and north</i> (Story 5)	information object (about)	<i>a book about Canadian travel writing in the North</i>	<ul style="list-style-type: none"> <li>information object (that is part of information object)</li> <li>information object (that is subject of an information object)</li> </ul>	<ul style="list-style-type: none"> <li><i>the chapter of the book</i></li> <li><i>the book that has been reviewed in the review article</i></li> </ul>
concept and person	<i>sociology and childhood and Giddens</i> (Story 9)	information object (about concept and person)	<i>the works of sociologists or LIS researchers who use Giddens' theory in their studies of childhood</i>		
person	<i>"Jennifer Egan" Bakke</i> (Story 4)	information object (about information object)	<i>reviews and critiques about the translated book</i>		
person	<i>Tom R. Burns</i> (Story 3)	person (as an entity)	<i>who is he? his background and his affiliation</i>	<ul style="list-style-type: none"> <li>information object (by person)</li> <li>person</li> </ul>	<ul style="list-style-type: none"> <li><i>his publications</i></li> <li><i>co-authors</i></li> </ul>
person	<i>Tom R. Burns</i> (Story 3)	information object (about information object)	<i>reviews about his article</i>		
person	<i>Elsa Schiaparelli</i> (Story 1)	physical man-made thing (by)	<i>visual of the garment</i>	<ul style="list-style-type: none"> <li>physical man-made thing (entity)</li> <li>legal body</li> <li>physical man-made thing</li> <li>activities</li> </ul>	<ul style="list-style-type: none"> <li><i>material and detail description of the garment</i></li> <li><i>where it is held?</i></li> <li><i>other objects created by her</i></li> <li><i>process of producing wool</i></li> </ul>

Continuation of Table 5.1					
Query		Primary need		Secondary need	
Entity	Example	Entity	Example	Entity	Example
person	<i>Elsa Schiaparelli</i> (Story 1)	context	<i>getting an idea</i>		
person	<i>Christian Berard</i> (Story 1)	person (as an entity)	<ul style="list-style-type: none"> <li>• <i>who is Christian Berard?</i></li> <li>• <i>what he looks like?</i></li> <li>• <i>his biography</i></li> </ul>	<ul style="list-style-type: none"> <li>• context (around a person and his activities)</li> <li>• person</li> </ul>	<ul style="list-style-type: none"> <li>• <i>about his collaboration with Schiaparelli and others</i></li> <li>• <i>If he was a minor person or a famous one</i></li> <li>• <i>Schiaparelli's other collaborators</i></li> </ul>
person	<i>Schiaparelli</i> (Story 1)	context (around activity)	<ul style="list-style-type: none"> <li>• <i>context around designing of the garment</i></li> <li>• <i>what was the fashion/ interest/ tendency in that month?</i> <ul style="list-style-type: none"> <li>• <i>was she the only one who designed this type of garment?</i></li> </ul> </li> </ul>		
person	<i>Richard Harrington</i> (Story 5)	legal body	<i>copyright holder of the photograph</i>	<ul style="list-style-type: none"> <li>• contact point</li> <li>• person</li> </ul>	<ul style="list-style-type: none"> <li>• <i>contact information of the copyright holder</i></li> <li>• <i>name of the people in the photograph</i></li> </ul>
person	<i>Richard Harrington</i> (Story 5)	information object (about activity)	<i>all the textual documents related to his journey to get the context</i>		
person and information object	<i>Richard Harrington and face of the arctic</i> (Story 5)	<ul style="list-style-type: none"> <li>• information object (about information object)</li> <li>• information object (about person)</li> <li>• context (around information object)</li> </ul>	<ul style="list-style-type: none"> <li>• <i>what people have said about him?</i></li> <li>• <i>what people have said about his travel book?</i></li> <li>• <i>how his book has been used as a source in other domains?</i></li> </ul>		

Continuation of Table 5.1					
Query		Primary need		Secondary need	
Entity	Example	Entity	Example	Entity	Example
person and information object	<i>David Vann and Jord</i> (Story 4)	<ul style="list-style-type: none"> <li>information object (about information object)</li> <li>information object (about person)</li> </ul>	<ul style="list-style-type: none"> <li><i>reviews and critiques about the translated book</i></li> <li><i>what they say about the translator?</i></li> </ul>		
information object and person	<i>“Gender and power” Connel</i> (Story 7)	information object (by)	<i>article by him</i>		

As Table 5.1 demonstrates, this study supports the finding that what users express as information needs in their queries does not necessarily describe their actual information needs. This corresponds with Ingwersen's (1992, p. 227) description of the label effect: "A conceptual 'distance' often exists between an information need, as represented in the actual user's mind, and the user's request formulation(s). Requests may consequently take the form of labels." Sometimes users' queries are very different from their actual needs, making it difficult to determine their actual information needs merely by looking at their queries. For example, in a query for David Vann and Jord in Story 4, the author and his translated book are the participant's expressed compromised need, but the participant's need is neither to find anything on David Vann, nor the translated version of the book themselves, but rather a review article on the translated version to find critiques about its translation.

According to Table 5.1, users gear their queries toward general and known entity types regardless of what their information needs are (Farrokhnia & Aalberg 2016), i.e., they still use general queries even when they search for something specific. The study participant from Story 5 provides a vivid example, in which a query is entered for "Richard Harrington", i.e., a person-entity type, to find some specific fact-based information (in this case, contact information for the museum) about a particular photo that Harrington took. Although she has specific details about this photo, she still conducts a general query – in this case, a person-entity type.

Table 5.1 also demonstrates that one query and its actual respective information needs (primary needs) also can imply several other actual information needs (secondary needs) that are interconnected. For instance, in Story 1's query on person-entity type, clothing designer Elsa Schiaparelli, the participant's information need is to find a picture of a garment that Schiaparelli designed. However, the story of the participant's information needs does not end upon finding the picture of the garment, as the participant realizes that she has several other information needs interconnected with the retrieved garment picture and Schiaparelli that need to be answered.

## 5.2 Types of information needs

In this section, the analytical results from users' information needs beyond the compromised level of Taylor's information-need levels are discussed. Users' information needs are grouped into four principal categories (Fig. 5.1), which grew iteratively from research-data analysis. The emergent categories comprise: 1. fact-finding; 2. resource-finding; 3. entity-finding; and 4. context-finding. In the information need-type continuum (Figs. 2.1 and 5.1), on the endpoint with the lowest level of the unknown, uncertainty and openness, the emergent types of fact-finding, part of resource-finding, and part of entity-finding can be characterized as Kelly's (1970) indicative mood, which is when the user knows what is necessary to satisfy the need and knows the answer, or at least a form of the answer or the resources that hold the answer. It also is at this endpoint of the continuum where Taylor's (1968) command type, Bates' (1998) known item, Ingwersen and Järvelin's (2005) specific type, Ingwersen's (1986) verificative, and

Marchionini's (2006) look-up tasks are located.

On the middle of the continuum (Figs. 2.1 and 5.1), part of resource-finding and part of entity-finding can be categorized, along with Ingwersen's (1986) conscious topical, Marchionini's (2006) networks of related concepts, and Cole et al.'s (2010) known-item search 2 and unknown-item search 1, in which the user possesses some measure of the unknown, uncertainty, and openness while simultaneously being open to finding different aspects and angles of a known or unknown thing.

The other end of the continuum - which represents the highest level of the unknown, uncertainty, and openness, when the user does not know what can satisfy his or her information need and is open to new ideas - can place the context-finding type alongside Taylor's (1968) question type and Ingwersen's (1986) muddled-topical type. The full explanation for each category is presented below. Each category will be clarified further through some corresponding real-world examples from the research data.



Figure 5.1: Cultural heritage information needs of users

### 5.2.1 Fact-Finding

Fact-finding is a type of information need that corresponds to the same type found in extant research, e.g., Hennieke (2017) and Duff and Johnson (2001). In the fact-finding type, the user is looking for a specific bit of information, differentiating it from other types. It differs from the entity-finding type, which includes a range of factual information, along with other related structured information, and is not as limited and specific as the fact-finding type. The response that can be expected to satisfy a fact-finding type of need would be specific factual information related to any types of entities, such as a person, thing, event, place, or time – more specifically, the name of a particular person, inanimate object, occurrence, geographical location, date, or time period, etc. Below are extracts from interviews to provide solid examples of fact-finding information-need types.

**Contact information and organization name** One example of fact-finding type in the present study's research findings can be found through the participant in Story 5, who was looking for the name and contact information of a photo's copyright holder to seek permission to use the photo in her book. Another example is the participant in Story 1, who wondered in which

museum or gallery a particular garment can be found, either to visit the museum in person or call to ask for more information. In both of these examples, participants searched for particular facts, i.e., the name of an archive, museum, or gallery and an email address or phone number of that entity. In these cases, the users' information needs are of the fact-finding type, though it might seem like these needs are of the resource-finding type because to find these fact(s), they first need to find the objects' images (i.e., the photo or the garment). However, through finding the resources, they hope to be guided toward the factual information that they seek. For example, within the meta-data description provided for the object(s), and following existing, related links, they hopefully can find this fact-based information (in this case, contact point) that they seek.

**Name of person(s)** In another example, the participant in Story 5 wishes to find the name of the people in a particular picture to use as a footnote in her book. The participant seeks factual information – in this case, the names of particular people. Another example is the participant in Story 2, who submitted a query on a rural-employment guarantee law as a particular concept and specifically was looking for people and organizations that have written something about the law or have been involved in its formation. Thus, this search aimed to find the names of people or organizations.

### 5.2.2 Resource-Finding

The resource-finding type of information needs comprises information needs in which the user seeks information resources and materials that carry related information to what the user wants. It can be either a particular information resource or any kind of information resource(s) generally associated with a concept, person, event, or thing about which the user wants to know more. Cultural-heritage information resources in archives, libraries, and museums vary, from scientific journals or books, to paintings, garments, newspaper articles, archaeological reports, or letters. This type of information need corresponds with Hennieke's (2017) resource discovery questions including material-finding, specific type, specific item, research questions, and consultation. The resource-finding type of information need also relates to a combination of Duff and Johnson's (2001) material-finding, known item, specific form, consultation, and education types of questions asked in archives. With this type of information need, i.e., resource-finding, users are searching for resource(s), resource parts, or associated resources:

#### Resource as a whole

In this category, the user wants to see the entire information resource to learn about its content. For example, the user may want to see an entire article or the image of a museum object to learn what has been written about the object or what it looks like. One example from the present study's findings is an interview participant who wanted to see a visual representation of a garment that originally was designed by Elsa Schiaparelli, but now has been counterfeited and has been the subject of a trial. The participant wants to find the garment or a picture of it to see what

it actually looks like. Other examples in this category include the participant searching for a book about Canadian travel writing; the participant who wanted to find some articles about the evolution of social system, particularly to see what kinds of discourse exist on this concept; the participant who has found a book entitled *Iron age Myth and Materiality: an Archaeology of Scandinavia AD 400-1000*, by Paul James Cowie, and wanted to search for other works that Cowie has written on the topic of materiality in archaeology; and the participant who wanted to find all publications by Kurt Iveson, her favorite author, whose research interests resemble hers. In all these examples, the participants are looking for unknown information resource(s) that are about a particular concept and/or created by a known person. However, sometimes users are looking for a known information resource. Examples from the findings include the participant in Story 7, who wanted to find a particular article, titled “Gender and Power”, by R. Connell, and the participant in Story 4, who has a book about actor-network theory in Spanish and was looking for an English-language version.

### **Resource parts**

Sometimes when the user looks for a resource, he or she does not want to access the whole resource, but is interested only in specific parts of it. For example, in research articles, users’ principal intention may be to see only the methodology and sampling section, the theoretical approach, or graphical elements, e.g., a flowchart or diagram, but not the whole article (as their primary intention). An example of this subcategory is the participant in Story 6, who is at an early stage of his project and is looking for articles about similar projects on value chain in medical plants to scrutinize his methodological approach. Thus, he is particularly interested in finding research articles’ methodology sections. Another example of this subcategory is the same participant in Story 6, who would like to find a graphical representation of the concept, such as a flowchart that represents the concept of value chain. The participant is not that interested in reading the whole article. The principal purpose of his search is to find a graphical representation of the concept. It is worth noting that in this example, the participant may take a look at the article’s text later if something in the flowchart needs clarification.

### **Associated resources**

In other cases, the user is searching for a resource, but he or she is not that interested in the principal resource initially, but only particular resources associated with it, e.g., an image, building plan, interview guide, or survey questions, among other materials. Such associated parts sometimes are found in the resource’s appendices. For example, through searching for a particular concept or author, a participant who is designing the methodology of his research may be looking for research materials such as an interview guide or survey questions. Thus, he actually is more interested in research material such as interview sheets related to the methodological approach than the article’s principal body of text.

### 5.2.3 Entity-Finding

Another type of cultural-heritage information need is entity-finding, in which the user wants to find all related information about a particular thing, concept, person, event, etc. This type of information need is common when the user encounters something new and unfamiliar, and wants to figure out what it is. This can happen either at the beginning of a search task or in the middle, when the user encounters something new. Information that can satisfy this type of need can be found from a well-structured information source, like a Wikipedia page, in which the user can get an overview of all information related to the subject. This type of information needs is in line with what IR experts study at TREC (Text REtrieval Conference) Entity track and The Initiative for Evaluation of XML Retrieval (INEX) Entity Ranking track on related entity finding and entity retrieval (De Vries et al. 2007, Balog et al. 2010, Balog 2017).

The information that users discover during entity-finding pursuits is important in leading them to the next steps in their research to help them determine their research direction. The following examples from interviews represent the entity-finding type of information needs, in which participants are searching for information about a particular concept, person, or thing.

#### Concept as an entity

The entity-finding type of information need can be about a particular concept, in which a user wants to know the definition of a concept, its assorted aspects, its use, etc. For example, the participant in Story 2, who is in the early stages of a research project may be looking for any and all information about the Rural Employment Guarantee Act, especially details on its formation and implementation, those who participated in creating and passing it, any other stakeholders tied to the act, when it was enacted, how it is perceived and functioning, etc. Another example of the entity-finding type is a participant in Story 6, who is searching for value chain as a particular concept and wants to find everything about this concept. He wants to know what a value chain is, including the concept behind it and other aspects associated with it.

#### Person as an entity

The entity-finding type also can be about a particular person. This usually happens when the user encounters a new name – such as a new, unfamiliar author – during a search. To continue his or her principal task, the user wants to get to know this particular person who is unfamiliar by searching for all information directly related to that person. This can include facts about the person, such as nationality; language; birth/death dates; family; profession; community, professional, or other group affiliations; or works that he or she has created (e.g., published articles, paintings). From the present study's findings, one example is the participant in Story 3, who is looking for all available information about Tom R. Burns, including his general biography, affiliations, academic background, works, co-authors, and other, more general items. Another example is the participant in Story 1, who was looking for information about the fashion designer, Schiaparelli and finds a letter written by the designer to a person named Christian Berard. The

participant becomes curious and wants to know more about this person and how he relates to this designer. First, the participant wants to find a picture and overall biography, including details on occupations, what he has accomplished, whom he has worked with, etc. In other words, the participant is interested in getting an overview of this particular person.

### **Thing (physical item/conceptual object) as entity**

Sometimes entity-finding is about a particular thing, which can be physical or conceptual. The kind of information that meets users' information needs in this type of search would be everything about that particular thing and can vary from information about the physical description of the thing to all related activities and events that occurred during the thing's existence, e.g., its creation, modifications, acquisition, production, transfer of custody, relocation, etc. An example of this is the participant in Story 1, who wanted to know everything about a particular winter sweater designed by popular French designer Schiaparelli. The participant sought entity-based information about this particular sweater's physical attributes, including general visual appearance, textile/material, dimensions, color, design, name, intended purpose, physical location, and different activities associated with its design/creation, production, etc.

### **5.2.4 Context-Finding**

Context-finding is another type of information needs introduced here, in which the user usually is not strict about finding specific kinds of information, but is completely open to finding any and all kinds of information that provide contextual data for the user's initial information need. Thus, with this type, it is not easy to identify a specific type of information that can satisfy the user. It can vary, from a scholarly journal article or archival document, to an irrelevant ad or a picture in a magazine. Anything, even if not directly related to the user's initial need, can provide insights and perspectives on the user's information need to cultivate a better understanding of it. The following sections provide real-world examples:

#### **Context around thing (physical item/conceptual object)**

In the context-finding type of information need, a user sometimes wants to find contextual information about a particular thing that can be either a physical object, or a conceptual object, like a painting, article, design, etc. One aforementioned example is the participant in Story 1, who was searching for a particular garment designed by French fashion designer Schiaparelli. After finding an image of this garment and some information that mostly described its appearance, the participant wanted to find context information around it, e.g., how the garment was produced, who provided the wool needed to knit it, and whether knitting was a popular activity among the women of that time, etc. Moreover, the participant wanted to find the specific museum collection in which this particular garment was being exhibited to examine other objects in the collection, as well as other creations by this designer. Also, she wanted to know how many copies of this garment exist in different museums to help determine whether it was a popular garment at that

time. Another example is the participant in Story 3, who was looking for any discussion and critiques about Tom R. Burns' article. In this case, the context around this particular article is important to the participant.

### **Context around a person**

Users in some cases are looking for contextual information about a particular person. An example is the participant in Story 4 who wanted to find a discussion and critiques about a known translator. She is interested in finding out what people have said about this translator to get the context around him and his profession as a translator. Another example is the participant in Story 1 who is looking for an idea and would like to find any contextual information about the designer and her professional life.

### **Context around an event/activity**

Sometimes users are looking for contextual information around a particular activity or event. One example is the participant in Story 5, who wanted to find all textual documents to get more context related to Richard Harrington's journey. Such information could be any piece of information about Harrington travel's to the North, such as data in archival textual documents, including, but not limited to, his mission, how he procured the money for this journey, obstacles along the way, related correspondences, what happened before and after his journey, and everything about/around him and his travels. Another example is the participant in Story 1 and her searches about a particular sweater by French fashion designer Elsa Schiaparelli. After finding the picture of this garment in a fashion journal, the participant wanted to find contextual information about how this garment was designed. For example, to get the context, she said she wanted to know whether the designer became famous because of this design; what the editor of this journal said about the designer, whether the editor meant to make the designer famous by introducing this particular design in the journal, what the fashion trends were at that time, whether the designer had been the only one to create this type of garment, what the topic of preoccupation was like, and what people thought about fashion at that time, including their preferences, etc. Such information provides the participant with context to give her some insights about this designer's creation of this garment.

### **Context around a concept**

Concepts can be used and discussed in different domains and disciplines in different ways and from different perspectives. For example, it is important for the participant in Story 1 to see how a particular concept (in this case, intellectual property) has been used and discussed among scholars in the history and sociology disciplines. She wanted to know how this concept has been perceived in these disciplines and any research conducted on it. As a relevant result, the participant found a book about debates around the searched concept. In another example, the participant in Story 2 is preparing to begin the practical part of his research and is looking for

sample searches on a particular topic, human right and ecology. However, the participant is interested in some contextual information on works that have been done on this topic to get some idea about core researchers on the topic and those with whom he would need to interview as his research sample. For instance, the participant wanted to find the number of times people have been collaborating together on a work about the participant's chosen topic, or whether these people have worked (written) about a specific aspect of the topic and how many times they did so. In fact, he said that getting such contextual information can help him find the key people tied to different aspects of this particular topic.

**Contribution to Ingwersen's categories of information needs** The four emergent categories of information needs in this study can be viewed as a contribution to Ingwersen's three categories of information needs – verificative information need (VIN), conscious topical information need (CIN), and muddled topical information need (MIN) – by making them more specific and challenging them on some characteristics. The fact-finding type is in line with VIN, and the resource-finding type corresponds with CIN. With the entity-finding type, the user seeks some “new” and “unknown” concepts or things that may be completely “outside of known subject matter or domain”, like what Ingwersen described in MIN. However, users' level of uncertainty and exploration is not the same as the one in MIN. The user is not going to explore all the information in the information system, like what he or she does in an exploratory search. Instead, his or her entity-finding type of information need gets satisfied by finding a Wikipedia page on that unknown concept or thing to get an overview of the “unknown” concept or thing. Moreover, the context-finding type of information need is in one way similar to Ingwersen's CIN in the sense that it is about pursuing “different aspects” and angles of “known-subject matter”, but is not about “clarifying” or “reviewing” them. On the other hand, it is similar to MIN in the sense that the user wants to “explore” new concepts, but it differs from MIN because it is not necessarily about exploring concepts “outside of known subject matter or domain.”

### 5.3 General patterns of users' information needs

In this section, we aim to understand and formalize what users are looking for and conceptualize their information needs through an ontological perspective. As the result of data analysis in the last two phases of analysis, some general patterns in participants' information needs have been emergent. Participants' information needs have been interpreted based on their own words and modeled in CIDOC CRM and its domain-specific extensions. CIDOC CRM has been selected as an ontology to be used for ontological analysis in this study because it is the dominating reference model for cultural-heritage information. The researcher that the CIDOC CRM entities and properties inspire named the emergent patterns in users' information needs in this exploratory study. These general patterns are described in two categories based on the participants' real-world information needs' targets.

- Category 1: Three principal notions of actor, thing, or event as the target of information need
- Category 2: Related information on actor, thing, or event

The following section provides an ontological representation of the mentioned general patterns in users' information needs when they search for cultural-heritage information. The ontological modeling represents the foremost, simplified patterns backed by empirical evidence found in users' information needs in this study. Each pattern will be introduced with a description and a diagram showing its ontological structure, followed by some examples that have been represented thoroughly in the previous section. The bold, colorful borders around the CIDOC CRM classes represent the patterns' principal notions.

### 5.3.1 Category 1: Three principal notions of actor, thing, or event as the target of information need

Sometimes users specifically are looking for actor(s), thing(s), or event(s)/ activities. The three notions of actor(s), thing(s), or event(s)/ activities are among the CIDOC CRM model's principal notions. Based on this study's research data, actor(s) specifically can be the principal target of information need, in which participants look for actor(s) who:

- create or possess a particular thing (information object/physical man-made thing),
- work on a concept or activity together (colleagues/collaborators)
- are addressees of someone's correspondence.

Regarding a thing as the target of information needs, participants would like to find thing(s) (information object/physical man-made thing) that:

- are created by an actor
- are inspired by another thing
- refer to an actor, event, or thing
- are part of a collection or form part of another thing
- are a translation of another thing

Considering event/activity as the target of information needs, participants are interested in finding event(s)/activity(ies) related to a particular person or thing.

**Network of people working on a concept**

The pattern of the network of people working on a concept (Fig. 5.2) focuses on those who work on a particular concept through common or separate activities. Different people (E21 Person) carry out these activities (E7 Activity), creating a thing that can be either an information object (E73) or a physical man-made thing (E24). The created thing is about a particular concept (F6) that is of interest to the user. This pattern covers that part of users' cultural-heritage information need in which they want to know who has worked on a particular concept (F6). The pattern focuses on people (E21) specifically, not the information objects (E73) that they create. An example of this pattern is the participant in Story 2 who wants to know the people who worked on a particular concept, i.e., the Rural Employment Guarantee Act, to find interview prospects. Thus, the people (E21 Person) are the principal target of the participant's information need.

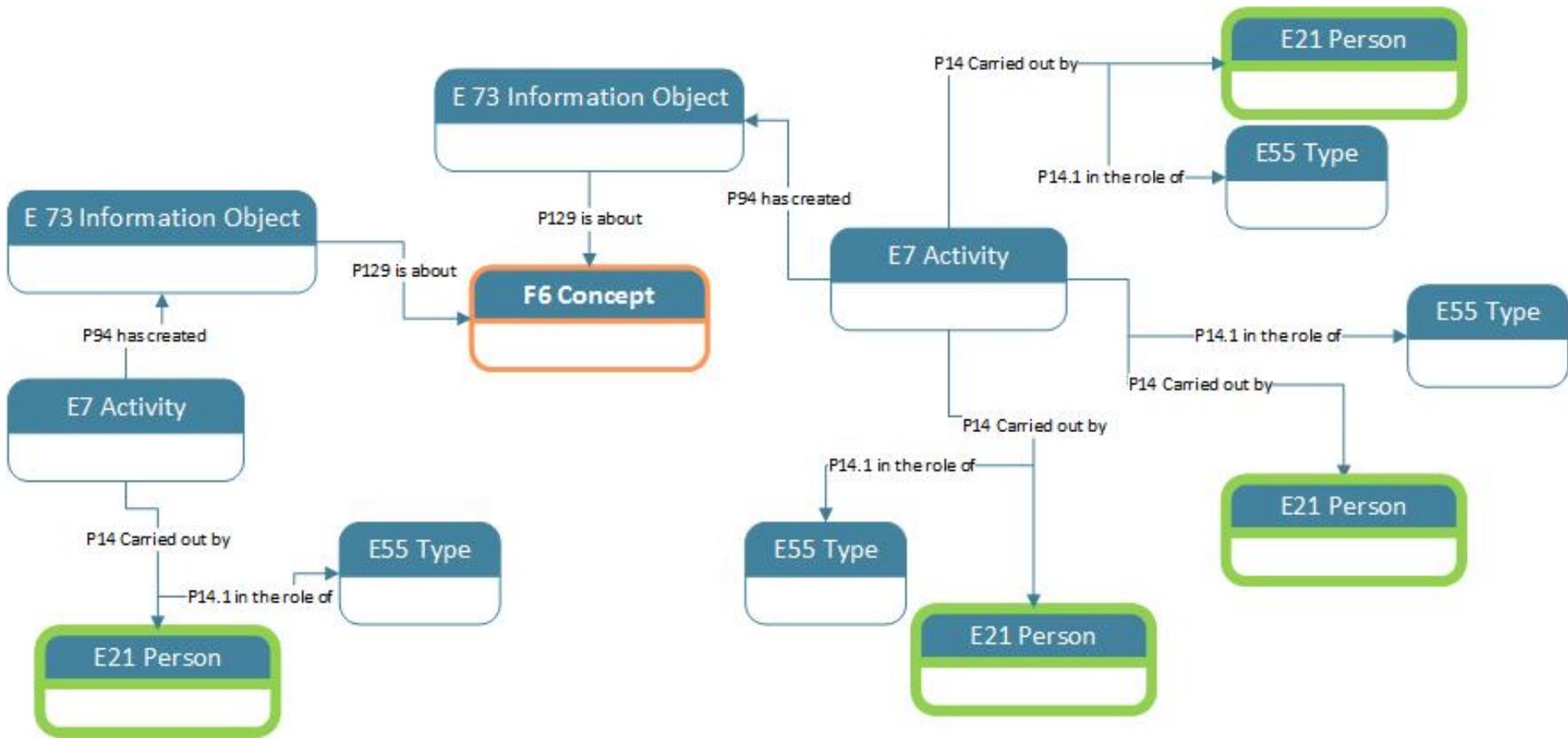


Figure 5.2: Network of persons working on a concept

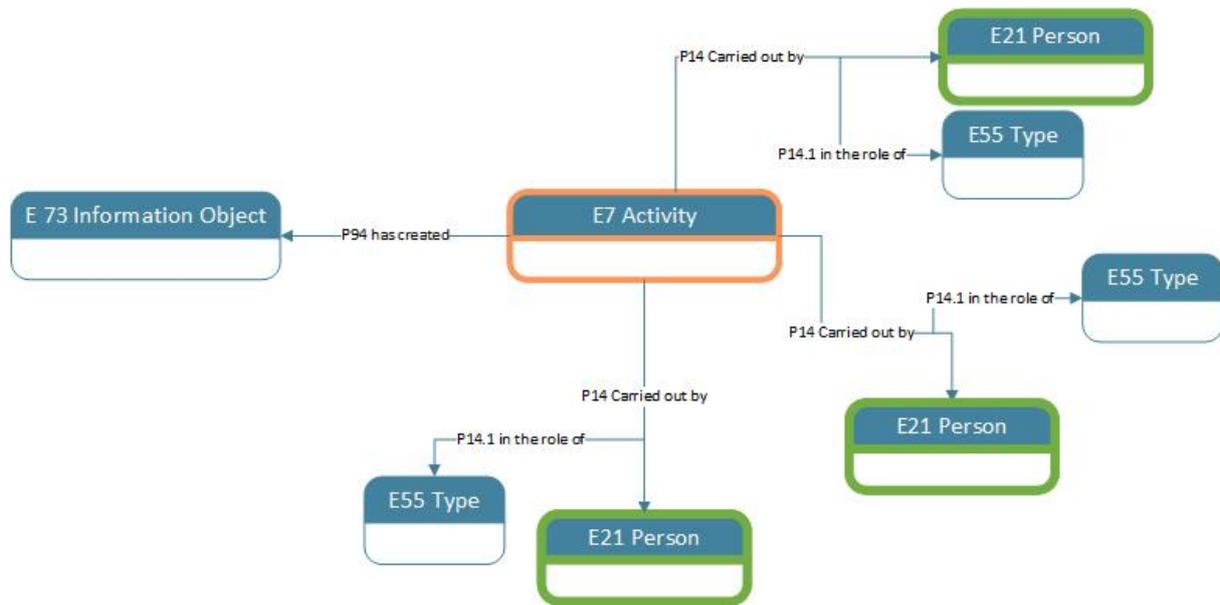


Figure 5.3: Network of collaborators on an activity

### Network of collaborators on an activity

The network of collaborators' pattern in a particular activity (Fig. 5.3) covers the context of collaboration, the collaborators, and the role that each person plays in this collaboration. The core class in this pattern is a particular activity (E7) carried out (p14) by different people (E21). In this pattern, each collaborator (E21 Person) can participate in the activity in the role of (p14.1) any type (E55). The target of the user's information need in this pattern specifically is people (E21) collaborating on a particular activity (E7). An example of this pattern is the participant in Story 2 who wanted to find people (E21) who have been involved in creating (E7) a particular law.

### Person's collaborators

The pattern of a person's collaborators (Fig. 5.4) focuses on the people (E21 Person) with whom a particular person (E21) shares activities (E7). These people may play different roles in the activity. The role that each person plays in performing an activity can be identified in the model through the notion of type (E55). This pattern answers to that part of information needs in which users want to find a particular person's professional network. An example of this pattern is the participant in Story 1 who looks for people (E21 Person) with whom Schiaparelli (E21 Person) has been collaborating during her professional life, as well as their collaboration types.

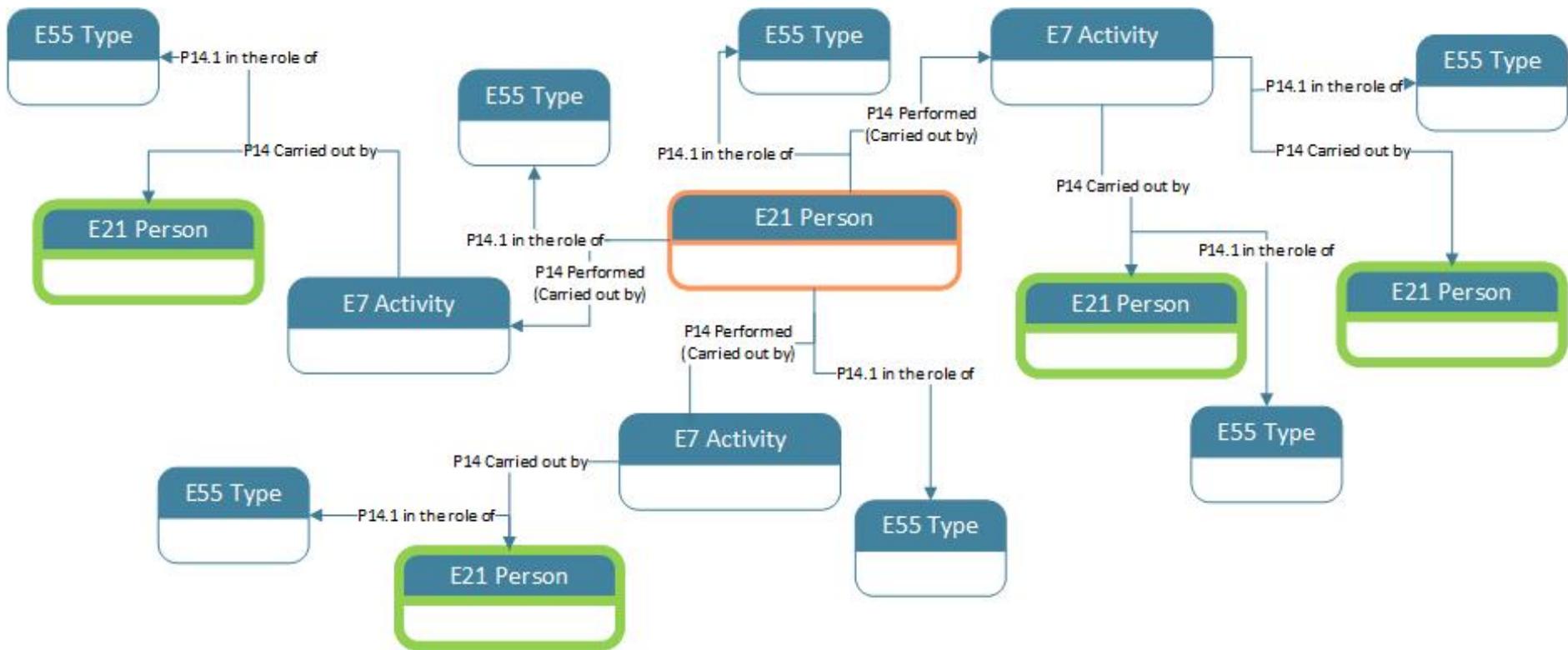


Figure 5.4: Person's collaborators

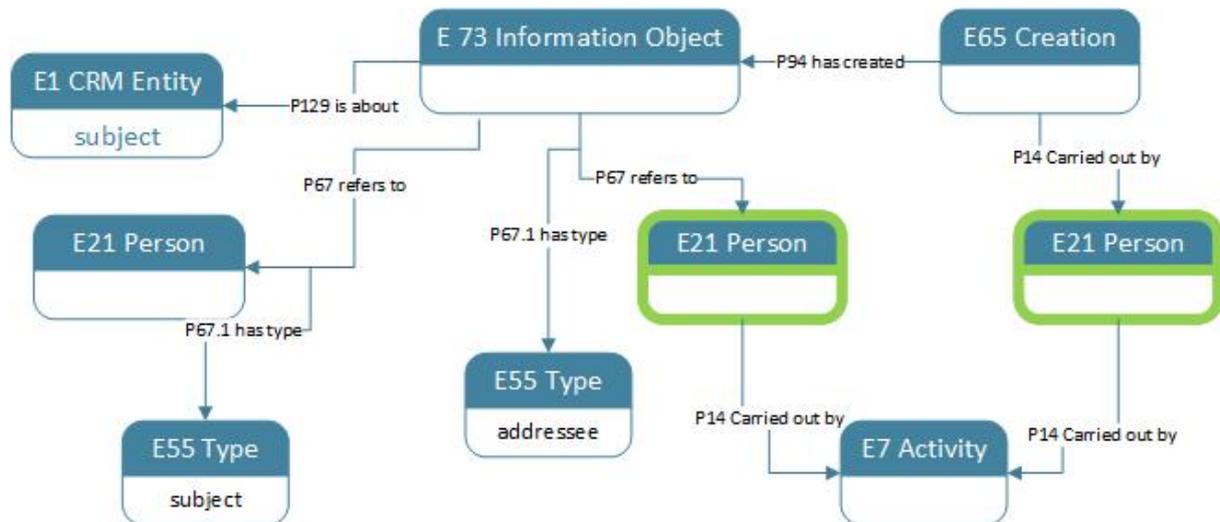


Figure 5.5: Addressee of a person's correspondences

### Addressee of a person's correspondences

The pattern of an addressee of a person's correspondences (Fig. 5.5) covers the context of a person who writes (E56 Creation) a letter (E73 Information Object) to a particular person (E21) mentioned as the letter's addressee. The letter content can be about or refer to anything (E1 CRM Entity). Similar to this pattern, Hennicke (2017) has identified the correspondence pattern as one of the ones that he found in the study of archival users' reference questions. In his thesis, for this pattern's ontological representation, he proposed some new properties including *has addressee (is addressee of)*, *had specific designation (was specific designation of)*, and *had general designation (was general designation)*, in addition to the CIDOC CRM properties used for the pattern. His pattern intended to cover both instances of writing a letter with the intention of mailing it to someone and a speech that addresses an audience. However, in the present study, the addressee's pattern for a person's correspondence(s) has been presented by the existing CIDOC CRM classes, properties, and sub-properties. The addressee pattern of a person's correspondences also can provide a leading point for the user to find the person's collaborators as well. An example of this pattern is the participant in Story 1 who finds a letter written by designer Schiaparelli and wanted to identify this letter's recipient, as well as any collaborations between them. In other words, the participant looked for the addressee (E21 Person) of the correspondence (E73 Information Object) and the letter's content (p129 is about).

### Works created by a person

The pattern of the works that a person creates (Fig. 5.6) focuses on the person's (E21) created entities (E73 Information Object or E24 Physical Man-Made Object) that have been created as the result of a creation activity (E65 Creation). A person (E21) can carry out (p14 perform) different creation activities and create different things, including propositional objects/information objects (E73) or physical man-made objects (E24). This pattern can answer information needs, such as all publications (E73 Information Object) that a particular author (E21 person) wrote (Story 3),

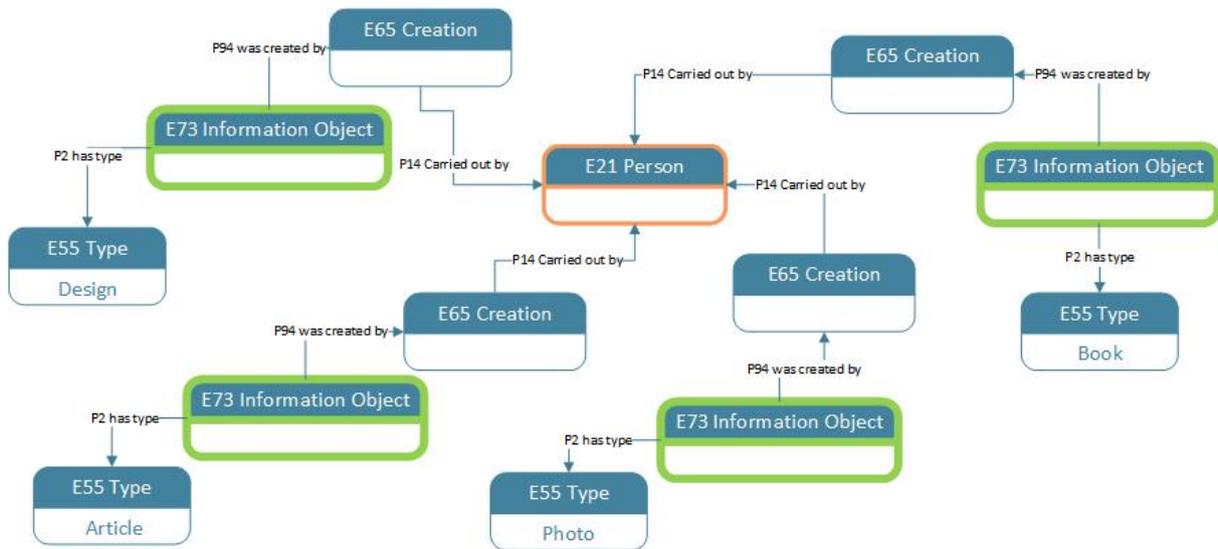


Figure 5.6: Works created by a person

the things (E73 Information Object) that a particular designer (E21 Person) created (Story 1), the photos (E73 Information Object) that a specific travel writer (E21 Person) has taken (Story 5), and the books that a particular person has authored (Story 7). The books, photos, articles, and designs are considered information objects (E73) that have been created through a creation activity (E65 Creation) carried out by that particular person (E21).

### Works inspired by information object

The pattern of works inspired by an information object (Fig. 5.7) covers the context of inspiration by others and their works in creating a new thing. In this pattern, during the process of creating a new work (E73 Information Object or E24 Physical Man-Made Object), the person (E21) carries out a creation activity (E65 Creation) in which the person (E21) is influenced by another work (E73 Information Object or E24 Physical Man-Made Object) that another person (E21) has created. An example is the participant in Story 9, who wanted to find research articles that used a particular theory (E73 Information Object).

### Discussions around a person and/or person's created work

The pattern of discussions around a person and/or a person's created work (Fig. 5.8) focuses on the discussions that other people have about a particular person and/or what this person has done. The discussions generally are not only about the person, but also implicitly about the activity (E65) that this person has performed and the work that this person has created as a result of that activity (E65). The work can be either a physical, man-made thing (E24) or an information object (E73). This pattern covers that part of users' information needs in which they want to know what has been said about a particular person (E21), his or her activity (E7), and his or her created work (E24 Physical Man-Made Thing) or (E73 Information Object). This information helps the user know the person and his or her work better. The information objects that can answer this kind

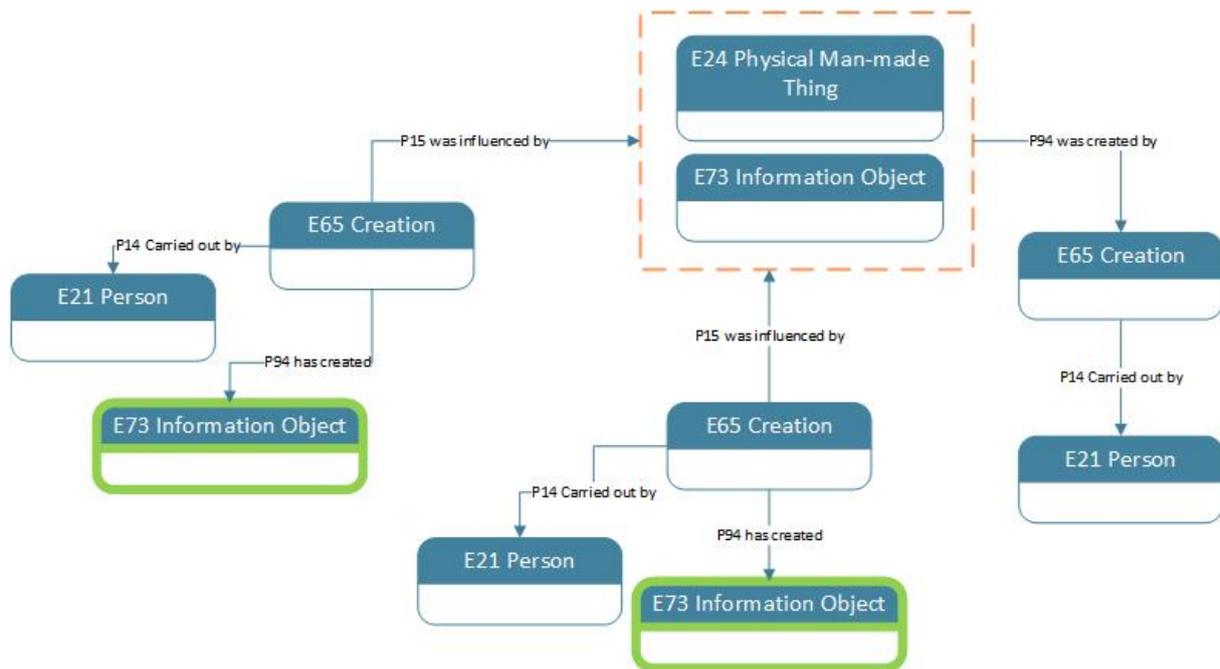


Figure 5.7: Works inspired by information object

of need can be of different types (E55), e.g., a review article (E55 Type), a critique (E55 Type) written in a newspaper, a discussion or interview (E55 Type) on a TV program, etc. An example of this pattern includes the participant who wants to know about discussions around a particular designer (E21 Person) and her specific design (E73 Information Object) in Story 1; discussions about a particular translator (E21 Person), his translated book (E73 Information Object), and, implicitly, the translation of the book in Story 4, or the discussions about a particular book or article (E73 Information Object) by an author (E21 Person) in Stories 3 and 4.

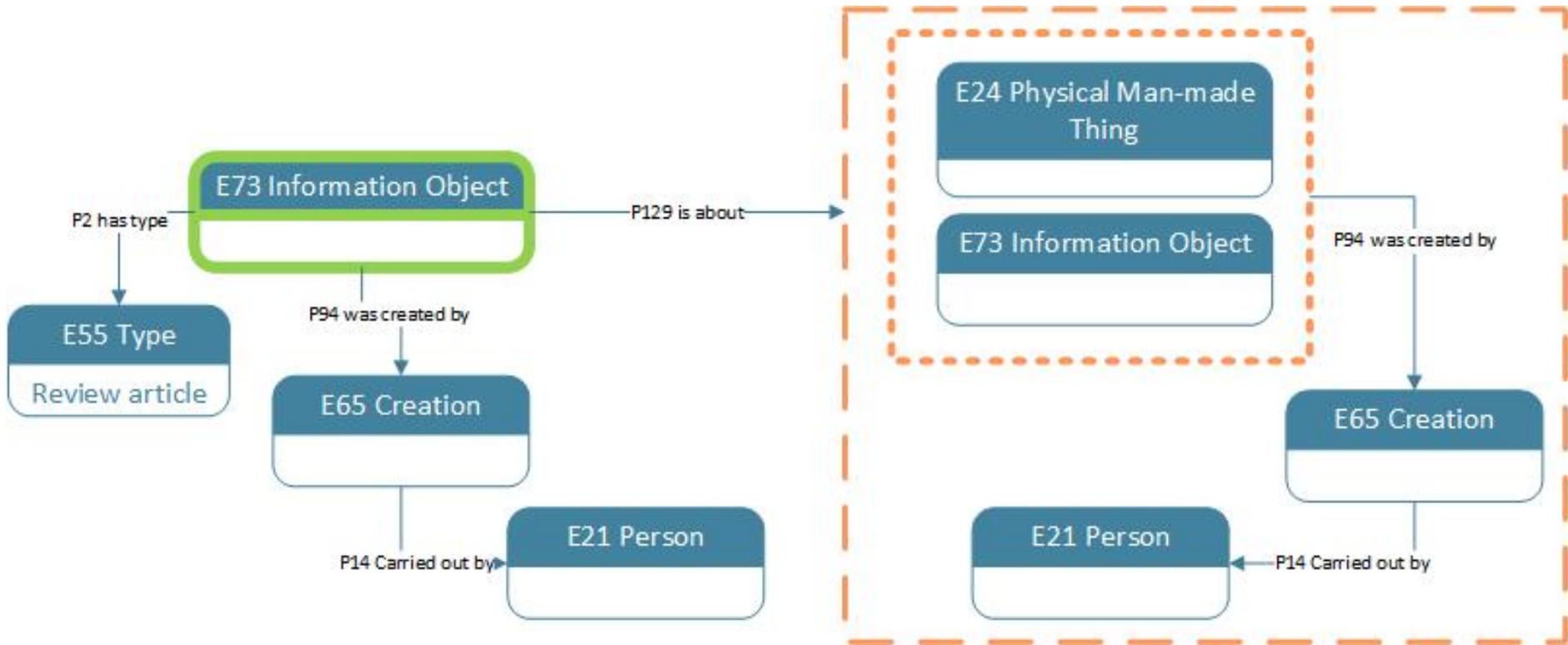


Figure 5.8: Discussions around a person and/ or person’s created work

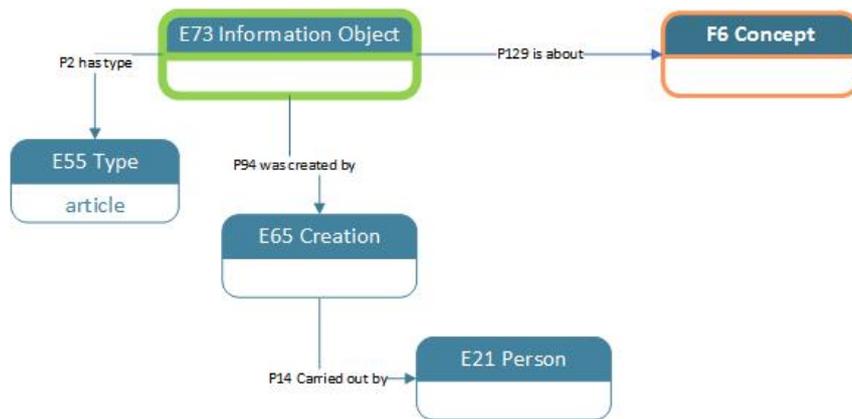


Figure 5.9: Discussions around a concept

### Discussions around a concept

The pattern of discussions around a concept (Fig. 5.9) focuses on a particular concept. Concepts can be understood and perceived differently by different communities. This pattern covers the part of cultural-heritage information need in which the user would like to know how a particular concept has been discussed in his or her field or other fields. The information objects (E73) that carry such discussions can be, for instance, a journal article, book, or newspaper article (E73). This pattern can be seen in the information need of the Story 8 participant, who wanted to see how the concept of materiality (F6) has been used and discussed in archaeology.

### Discussions around an activity

The discussion's pattern around an activity (Fig. 5.10) expresses an information need in which the user looks for any kind of information that relates in some way to a particular activity (E7) that a particular person (E21) or group of people carry out. An example of this pattern is the participant in Story 5, who wanted to find all the existing textual information about Richard Harrington's travels to the North (E7 Activity). In this example, the participant's desired information, as she mentioned, can be found in different types of information objects (E73), such as a government funding document, a letter of permission to visit specific places, the travel book that Harrington wrote, etc.

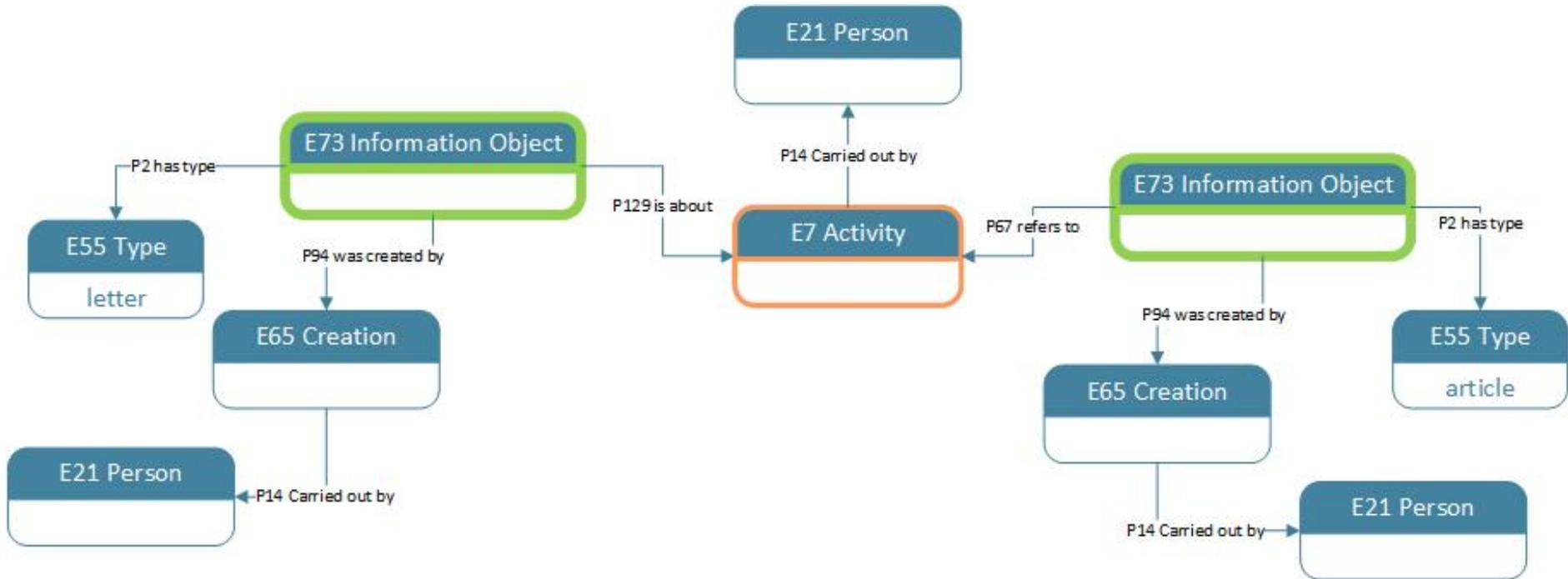


Figure 5.10: Discussions around an activity

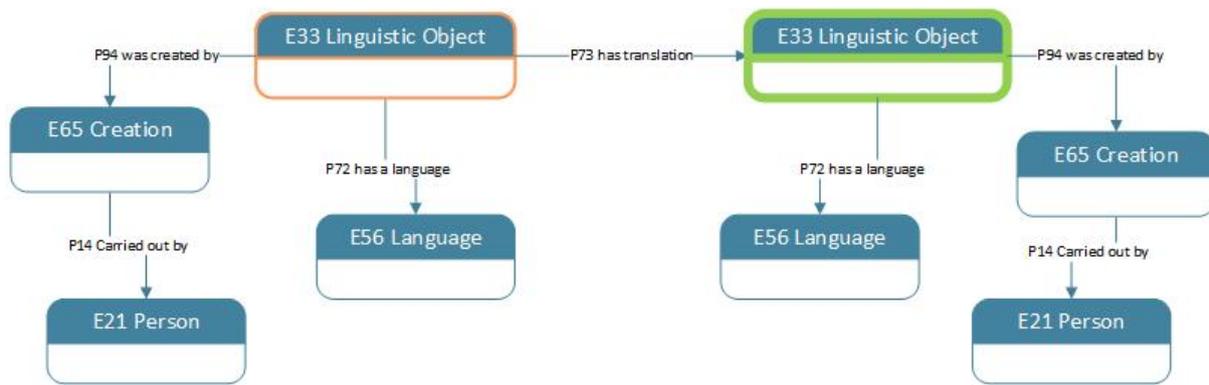


Figure 5.11: Translation of a work

### Translation of a work

The pattern of a work's translation (Fig. 5.11) focuses on the concept of linguistic objects as identifiable expressions in natural languages. A linguistic object in CRM is treated independently from its expressed medium or carrier. This pattern covers a written text that has been translated to another language. An example of this pattern is the participant in Story 4, who has a Spanish version of the book *Reassembling the Social: an Introduction to Actor Network Theory*, but starts searching for an English-language version.

In this study, we modelled the pattern of work's translation using CIDOC CRM classes and properties. However, it also can be modelled differently using FRBRoo classes and properties (see Aalberg et al. 2015).

### Event/activity related to a person or thing

The event/activity pattern related to a person (Fig. 5.12) or thing is an implied pattern in the user's information needs and covers the context of recorded activities carried out and events that happened during a person or thing's lifetime. A thing (E70) can be either a physical man-made object (E24) or an information object (E73), but the research data in this exploratory study do not support the event/activity related to an information object (E73).

Regarding an event or activity related to a person (E21), the pattern serves information needs about people's recorded personal life activities or professional activities that they have performed, or events in which they have participated. An example is the participant in Story 3 who wanted to find all research activities that a newfound author had conducted.

Regarding an event or activity related to a thing (E70), this pattern answers the information needs on the creation and production of a thing and all the events that happen or activities being carried out on this particular thing in a specific time frame. The activities related to a particular thing (E70) can be, for example, its creation, modification, transfer of custody, etc. An example of this pattern is the participant in Story 1, who discovered a new aspect of her research by discovering the activity of wool production (E7) related to producing a sweater in which she was interested.

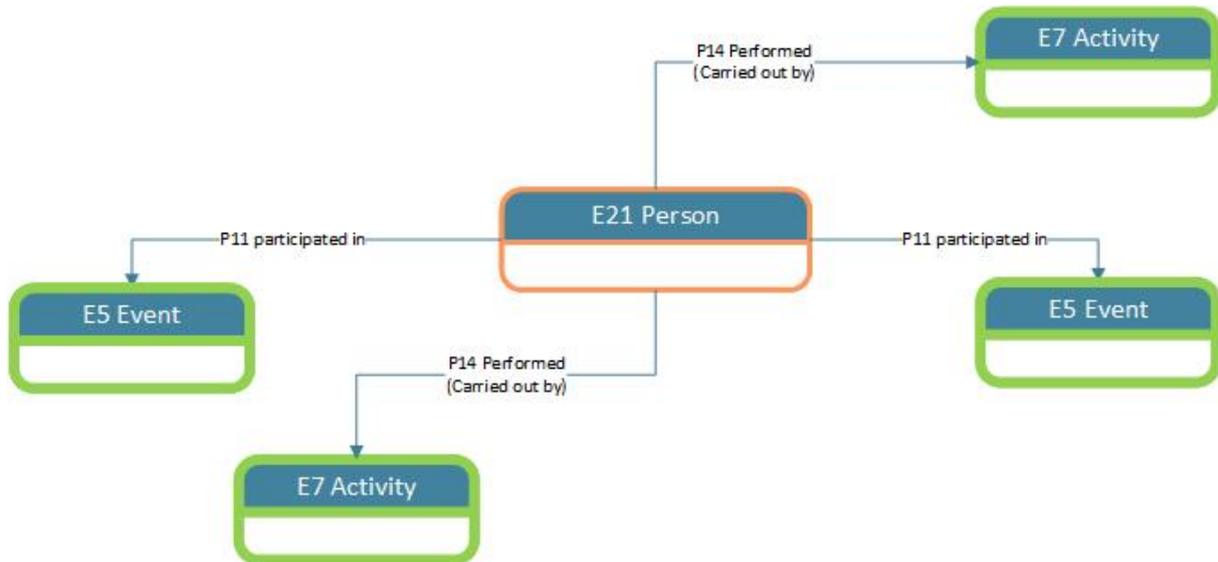


Figure 5.12: Events and activities related to a person

The target of information needs in this pattern would be activities (E7) or events (E5) related to a particular person or thing. This pattern can provide new insights to users by presenting different activities related to a person or thing and can help them find new aspects of their research.

### 5.3.2 Category 2: Related information on actor, thing, or event

Sometimes users are interested in finding fact-based information about the three aforementioned principal notions of actor, thing, and event/activity. In this category's ontological representation, desired information is found that usually is directly related to the principal notions. This fact-based information includes, but is not limited to, entity types such as appellation, title, parent, birth, death, time, place, dimension, language, creation, etc. This pattern also can provide new ideas by presenting different kinds of related information on what users are searching for.

Regarding factual information about an activity, one example is the participant in Story 5, who wanted to find all related information on Harrington's travels to the North (E7 Activity). Whatever was related to this particular activity – such as people, documents, or references to time and place, or other related activities – can elicit interest. An example of need for factual information about a thing is the participant from Story 1, who is interested in finding all related information about a particular sweater (thing). The desired information varies from detailed information about the physical description of the sweater, e.g., its color, material, texture, design, and title (if any), to where it is exhibited and its uses. Regarding factual information about a person, examples are the participant in Story 1 and Story 3, who wanted to find all related information on a particular person found on search-results list. The kind of information that can satisfy participants' needs pertains to the person's picture, general biographical information, affiliations, created works, and other activities.

Examples of this category of patterns are presented in Fig. 5.13 for information related to a

physical man-made thing and Fig. 5.14 for information related to a person.

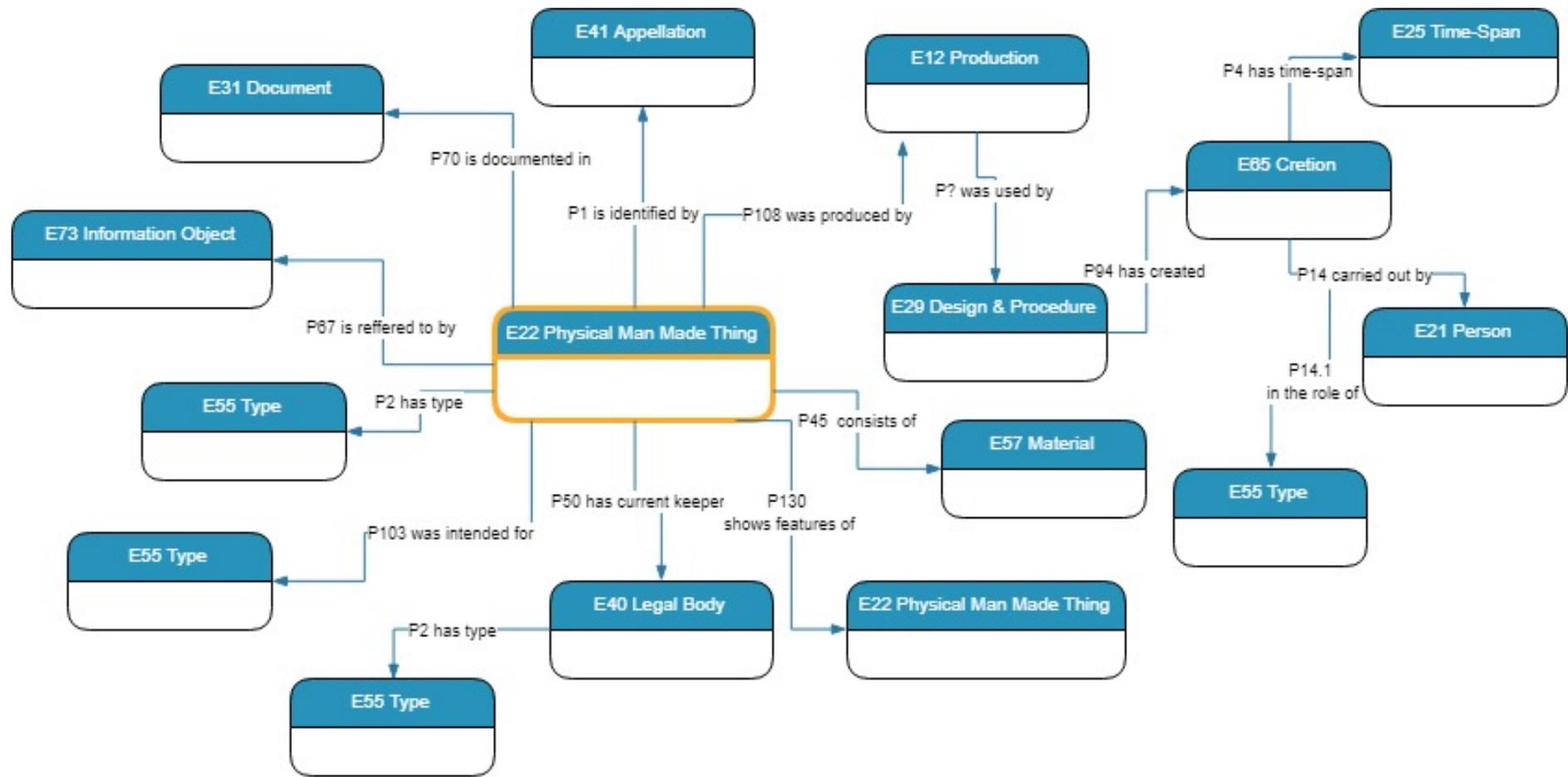


Figure 5.13: Information related to physical man-made thing



## 5.4 Model of Semantic Information Needs

The model of semantic information needs, shown in Fig. 5.15, is dynamic by nature and has been influenced by Taylor's (1968) information-need levels and the present study's empirical findings. Considering a complex task, a visceral need is a big and vague need that the complex task influences. This also corresponds with the first stage in the model of the information search process from Kuhlthau (1993a, 343):

At initiation [stage/phase], when a person first becomes aware of a lack of knowledge or understanding, feelings of uncertainty and apprehension are common. At this point, the task is merely to recognize a need for information. Thoughts are vague and ambiguous centring on the general problem or area of uncertainty.

Visceral need becomes a little bit clearer in the form of one or more conscious needs. At the next stage, the conscious need gets clearer and becomes a formalized need. In the end, the formalized need leads to one or more compromised needs, in which the user wants to express his or her need to an information system. During this development process, information needs become clearer, narrower, and more specific. At the same time, as a result of getting specific, the conscious need yielded from a visceral need does not include the visceral need entirely, but only the part that could be identified in the user's mind at any particular time. Thus, the rest of the visceral need, of which the user may or may not be aware, can be realized later based on the new information that the user gets during this time. Similarly, a formalized need does not necessarily include the entire conscious need, and compromised need may not express the formalized need entirely.

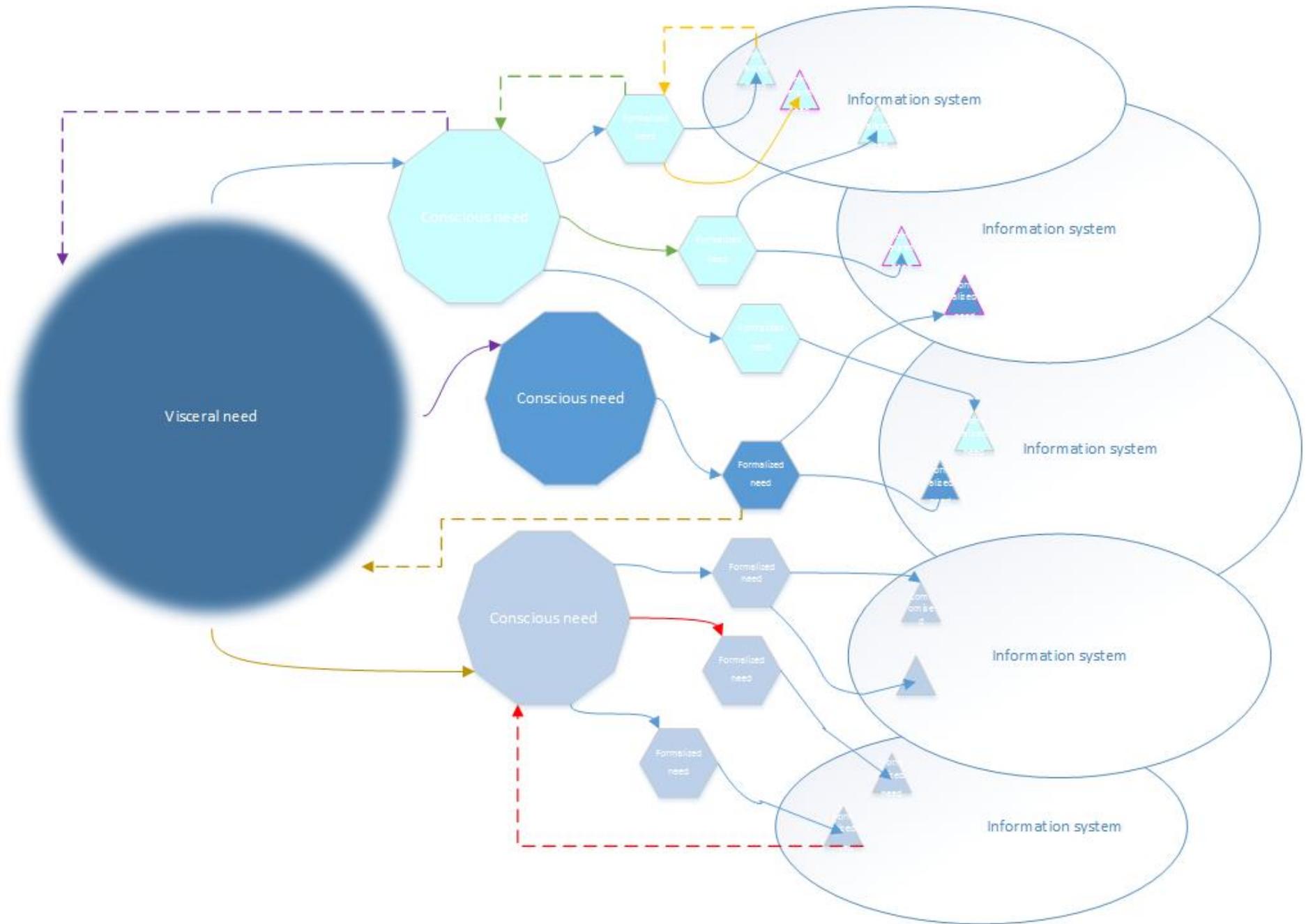


Figure 5.15: Model of Semantic Information Needs

The model of semantic information needs comprises the following principles:

- **Different potential conscious needs exist based on a particular visceral need.** They can be explained as different aspects of a visceral need.
- **Different potential formalized needs for a particular conscious need exist.** They can be different aspects of a particular conscious need that are expressed as different formalized needs.
- **Different potential compromised needs can stem from a particular formalized need.** Compromised needs are expressed in information systems as queries. Different compromised needs in the form of queries can follow one formalized need. Queries are refined several times, e.g., translated into another language, narrowed in meaning to be more specific, or broadened to be more general. Thus, these various queries try to answer one particular formalized need.
- **Different formalized needs can result in the same query type and content, although the root need is different.**

Such unique queries can be submitted to different information systems or even the same information system to look for different formalized and/or different conscious needs because a compromised need or query is an imposed expression of a user's need. The user may not have many choices due to different limitations. Accordingly, some of this study's findings are reflected in a Hofmann et al. (2009) statement regarding limitations in query analysis: "Any query can be interpreted in many different ways". As remarked in interviews and search episodes during this study, the same unique query can represent different information needs. For example, the participant in Story 1 queried the name of designer Elsa Schiaparelli for different information needs. She uses the designer's name as her query while searching for ideas, discussions about Schiaparelli's design, discussions around Schiaparelli's overall design activity, visual representations of one of Schiaparelli's designs, fashion, and a preoccupation at the time while she was designing the garment. In this example, one particular query can refer to different information needs that consider the dynamic group of information needs that includes a principal, vague, visceral need and different conscious and formalized needs:

- **The positions of different emergent needs vary in the Information Need Type Continuum (Fig. 2.1).** It means that each of the emergent needs in the model of semantic information needs is of different types regarding levels of the unknown, uncertainty, and openness.
- **Not all emergent information needs appear at the same time, but rather at different times and stages.** Users may realize their new information needs at any level based on how they are influenced by:

- the results from a compromised need, either satisfied or not satisfied with the results
- the mediator, which can be the information system and its features (e.g., through finding a new provider among the results category for providers)
- the process of information need getting realized in each of visceral, conscious, formalized, or compromised need levels

A compromised need, as a query, cannot express the entire conscious need, or even the entire formalized need from which it originated, i.e., we do not know what is going on in the user's mind with the process of information-need progression that starts at the visceral-need level and leads to the information need that is getting narrower and narrower at the compromised-need level. Thus, in studying users' information needs regarding their complex tasks, we cannot base our study on merely a single compromised need, or even a single formalized need. Information need should be considered a dynamic whole that includes a visceral need influenced by a complex task, different conscious needs rooted in the visceral need, formalized needs originating from conscious needs, and compromised needs that are imposed expressions of formalized needs. It also should be noted that some information needs end up at the compromised-need level and can be searched for in information systems, while with other information needs, it is not possible to search for them in databases and should be sought manually. The model of semantic information needs is, to some extent, in line with Cole's (2011) second proposition: "The user's information need does not evolve; only aspects of the topic [that] the user selects to investigate evolve or shift over the course of an information search". Within the model of semantic information needs lies a big, vague, visceral need that is concerned with a complex task, and searching for new aspects can be explained through users' realization of new conscious needs or new formalized needs regarding the same visceral need.

## Chapter 6

# Conclusion

The present exploratory qualitative study was conducted to investigate users' information needs when they search for cultural heritage information and determine how these needs can be mapped into an ontology, like CIDOC CRM, to provide more efficient information retrieval.

Following a naturalistic approach and contextual inquiry, we conducted contextual interviews with PhD candidates as active searchers to obtain information about their actual information needs regarding their PhD projects, which are considered complex work tasks. A practical method was proposed to analyze the qualitative data collected through contextual interviews using an ontological approach. Data have been analyzed through five different phases: thematic analysis and open coding, modeling the progression of information needs, creating the stories, ontological analysis, and final analysis (comparative analysis). One of the unique advantages of the proposed method to analyze the contextual interviews is that it provides the opportunity to apply a holistic approach to users' information needs that helps view these needs as a coherent whole, which its different parts are best understood as, in relation to other parts and within the context around them. The holistic approach that this study proposes helps us understand users' information needs by not only identifying and interpreting users' queries and their expected results related to their primary needs, but also considering their other related information needs, which we have termed secondary needs. Information-need studies based on queries ignore the holistic perspective on information needs, with attention concentrated on the limited compromised expression of needs, which does not necessarily represent actual needs. Information systems that incorporate the holistic view of information needs are challenged to consider not only users' primary needs, but also their secondary needs within the information-search process.

This study is among the first to investigate users' information needs holistically, considering both primary and secondary information needs, as well as compromised information needs (queries), through an ontology-based approach. In this study, we tried to answer the following research questions:

1. How do users search for cultural-heritage information?
2. What are users' expectations across information systems, and what types of information do they seek?
3. What common patterns can be discerned within users' information needs?
4. How can users' information-need patterns be projected onto CIDOC CRM and its domain extensions?

The specific objectives of this thesis were:

- To determine whether any similarities exist between ontology structures and researchers' information needs by exploring users' conceptualization of cultural-heritage information and their actual information needs.

- To identify characteristics and patterns in users' needs through ontological analysis of their actual information needs.
- To see whether and how well users' information needs can be projected onto ontologies such as CIDOC CRM and its domain-specific extensions.
- To model users' information needs through ontological representations that can be used in the application layers of future ontology-based information systems.

This study's findings showed that information inter-connectivity in cultural-heritage domains, reflected in ontology structures, also can be seen in people's information needs. The concept of relationships was shown to be an important part of how people search for cultural-heritage information and how they expect to satisfy their information needs. Users have an understanding that entities are related, which explains, for example, why they may search for a person, but actually want to find an object tied to that person. Faced with a complex search task, people look for different kinds of information. The present study's findings suggest that information needs can be grouped into four categories: fact-finding, entity-finding, resource-finding, and context-finding. They expect to satisfy not only their primary information needs, but also secondary needs, when they search for data in information systems. This study also proposes the concept of semantic information needs based on findings. The proposed model that presents the concept of semantic information needs is considered to be an extension of Taylor's levels of information need. The findings demonstrated that the common patterns of users' information needs that emerged from this study can be projected onto CIDOC CRM. This mapping of users' information needs onto CIDOC CRM showed that through a limited number of entities and properties, ontology-based information systems can satisfy different types of information needs.

## 6.1 Contributions

The contributions of this thesis are manifold, comprising theoretical, methodological, and practical contributions that are discussed in this section.

### 6.1.1 Contributions to theory

Regarding theoretical contributions, this study developed new concepts and categories, namely categories of users' actual information needs (Section 5.3), the holistic model of information-needs progression (Section 3.5.2), and the model of semantic information needs (Section 5.7), which can underpin development of more effective user-centered information retrieval.

**Categories of users' actual information needs** Traditional categories of information needs (previously discussed in Section 2.1.5) are either too vague for information systems to understand (i.e., categories based on users' cognition, with unknown or uncertain levels), or not

comprehensive enough (i.e., categories based on information's nature and characteristics) to cover all kinds of users' information needs in this exploratory study. Thus, a new categorization of information needs was needed. In this study, using an inductive approach, four information-need categories emerged from the empirical data: fact-finding, entity finding, resource-finding, and context finding. Considering the design of information systems, these categories provide an understanding of what kind of information should be returned to users to meet their actual information needs.

**Holistic Model of Information-Needs Progression** The holistic model of information-needs progression is an extension to Ingwersen's concept of polyrepresentation of the user's cognitive space. The holistic model of information-needs progression not only includes the concept of polyrepresentation of the user's cognitive space (Ingwersen 1996, p.41), representing the current information need, underlying problem space, actual work task or interest, and dominant work domain(s), but also the user's other related information needs that are connected semantically to current information needs and are realized during formation of current needs or as a result of the user's interaction with information systems and retrieved information regarding these current information needs.

**Model of Semantic Information Needs** Having a holistic approach to information needs in this exploratory study resulted in the emergence of the model of semantic information needs. The concept of semantic information needs is considered an original contribution to the conceptualization of users' information needs and to the development and refinement of contemporary information-retrieval assumptions and principles. The model of semantic information needs elegantly extends Taylor's levels of information need and provides a holistic framework demonstrating the interconnected nature of users' actual information needs in the context of complex tasks.

## **6.1.2 Methodological contributions**

### **From fragmented to comprehensive understanding of users' actual information needs**

The holistic approach to information needs used in this study provides a comprehensive understanding of users' information needs, which it views as a whole to be studied in context, i.e., from queries to actual primary needs to interconnected actual secondary needs. Both the holistic and ontology-based approaches to this study, along with the contextual interviews used as the method for collecting data and the holistic model of information-needs progression used as a tool to analyze the data, are significant methodological contributions to the field of information-need studies.

Furthermore, the emergent four categories of information needs, i.e., fact-finding, entity-finding, resource-finding, and context-finding, can be viewed as a method tool to analyze information

needs in future user studies.

### **6.1.3 Contributions to practice**

This study's emergent patterns and concepts are expected to hold practical implications for memory institutions, system developers, and users in the following areas:

#### **From single, stand-alone information needs to semantic information needs**

The data analysis revealed that users with complex tasks have different kinds of interconnected information needs, i.e., users' information needs should not be viewed necessarily as one single, stand-alone phenomenon, but rather as semantic information needs. Thus, the emergent concept of semantic information needs should be considered when developing future information systems to present cultural-heritage information in a way that satisfies users' semantic information needs, instead of only one single need.

#### **From meta-data-oriented information systems to entity-oriented information systems**

The emergent four categories of information needs revealed the need to move from meta-data-oriented information systems to entity-oriented information systems in which all entities (person, event, thing, ...) are interconnected to describe cultural-heritage information semantically. With traditional meta-data, the retrieved information is in the form of one or more records, each standing alone, but with entity-oriented information systems, users can find their desired information in the context of interrelated cultural-heritage information, in which pieces of information are connected to each other based on defined logical relationships. To provide such information systems to satisfy users' information needs, both memory institutions and system developers should consider entity-oriented approach in their future information organization strategies.

#### **From separate information systems to an integrated information system**

The concept of semantic information need emerged in this study as an important consideration for memory institutions and system developers to address some of the major limitations of current curating, indexing, and retrieving approaches in cultural-heritage domains. Users' different kinds of interconnected information needs can be satisfied through integrated information systems that provide different kinds of information and their relationships. This requires reconsideration of traditional approaches in organizing information based on information carriers and providers. Memory institutions and system developers should consider adjusting information systems to accommodate users' complex information needs in the Digital Age.

**From a complex ontology implication to feasible ontology implications**

General patterns that emerged from this study reveal that even though ontologies like CIDOC CRM include numerous entities and properties to provide detailed semantic descriptions of information bits, and its consequent complexity can create implementation challenges, ontology-based information systems can be built on limited types of entities and properties to meet a broad range of information needs.

**From query-oriented information systems to pattern-oriented information systems**

Findings from this exploratory study showed that queries can represent different kinds of information needs; thus, retrieving relevant information is not easy for information systems. However, the emergent patterns can help identify possible types of information that users are looking for, moving information systems closer to users' conceptualization of information. Thus, emergent patterns of information needs should be reflected in future information systems to support different kinds of needs in an integrated system to facilitate effective information retrieval.

**6.2 Future research**

Several promising directions have been forged for future research on conceptualizing users' information needs through ontological perspectives. This study's findings provide an important avenue to explore with detailed and rigorous extant studies. While this study demonstrated the importance of holistic approaches to information needs, the concepts and principles have yet to be explored, identified, and developed fully.

The general patterns and types of information needs that emerged from this study are in the context of complex tasks and expert researchers, so they need to be tested in other contexts as well.

The fact that context-finding was found to be one of the four types of users' information needs is an intriguing discovery that also should be examined more closely. By meeting the context-finding type of information need, information systems may intensify the problem of overwhelming the user when retrieving contextual information by presenting everything all at once. Further research into the context-finding type of information need should focus particularly on the most important entities and properties that need to be presented to provide the user with contextual information.

Another particularly promising area for future research is investigating the distance and paths between query entities and information-need entities.

The idea of providing richer indexing of information based on the entities and their relationships also is considered to be an important issue that is worth further exploration.

Furthermore, this kind of research requires a sufficient number of participants to make general comparisons. As this is an exploratory study, in which we attempted to explore users' information needs when they search for cultural-heritage information, a clear need exists for future studies so that the findings can be tested for empirical verification with larger and more targeted samples to clarify and amplify these findings and increase their generalizability.

Although the nature of this exploratory research precludes generalizations to other populations, it is hoped that the emerging concepts and patterns will shed light on the information needs of users with complex tasks. We also hope that our study will help encourage the Library and Information Science (LIS) community to conduct further research, using an ontology-based approach, on how to organize and present cultural-heritage information to meet users' semantic information needs.



# References

- Aalberg, T., Merčun, T., & Žumer, M. (2016). BIBSURF: discover bibliographic entities by searching for units of interest, ranking and filtering. In *Proceedings of the 16th acm/ieee-cs on joint conference on digital libraries* (pp. 207–208).
- Aalberg, T., Vennesland, A., & Farrokhnia, M. (2015). A pattern-based framework for best practice implementation of crm/frbroo. In *East european conference on advances in databases and information systems* (pp. 438–447).
- Agarwal, P. (2005). Ontological considerations in giscience. *International Journal of Geographical Information Science*, 19(5), 501–536.
- Andresen, L. (2006). Dublin core as a tool for interoperability: Common presentation of data from archives, libraries and museums. In *International conference on dublin core and metadata applications*.
- Aula, A., & Russell, D. M. (2009). Complex and exploratory web search. In *Information Seeking Support Systems Workshop (ISSS 2008), Chapel Hill, NC, USA*.
- Balog, K. (2017). Entity retrieval. In *Encyclopedia of database systems*.
- Balog, K., Serdyukov, P., & De Vries, A. P. (2010). *Overview of the TREC 2010 Entity Track* (Tech. Rep.). Norwegian University of Science and Technology Trondheim.
- Barr, A., Feigenbaum, E., & Roads, C. (1981). *The handbook of artificial intelligence, volume 1*. HeurisTech Press.
- Barry, C. L., & Schamber, L. (1998). Users' criteria for relevance evaluation: a cross-situational comparison. *Information processing & management*, 34(2-3), 219–236.
- Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online review*, 13(5), 407–424.
- Bates, M. J. (1998). Indexing and access for digital libraries and the internet: Human, database, and domain factors. *Journal of the Association for Information Science and Technology*, 49(13), 1185–1205.

- Bates, M. J. (2007). Invited paper What is browsing—really? A model drawing from behavioural science research. *IR information research*, 12(4).
- Bazeley, P. (2013). *Qualitative data analysis: Practical strategies*. Sage.
- Bekiari, C., Doerr, M., La Boeuf, P., & Riva, P. (Eds.). (2015). *Definition of FRBRoo: A conceptual model for bibliographic information in object-oriented formalism*. Den Haag: IFLA. [https://www.ifla.org/files/assets/cataloguing/FRBRoo/frbroo\\_v\\_2](https://www.ifla.org/files/assets/cataloguing/FRBRoo/frbroo_v_2). (Version 2.4. November 2015)
- Belkin, N. J. (1978). Information concepts for information science. *Journal of documentation*, 34(1), 55–85.
- Belkin, N. J. (1980). Anomalous states of knowledge as a basis for information retrieval. *Canadian journal of information science*, 5(1), 133–143.
- Belkin, N. J. (1990). The cognitive viewpoint in information science. *Journal of information science*, 16(1), 11–15.
- Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982). ASK for information retrieval: Part I. Background and theory. *Journal of documentation*, 38(2), 61–71.
- Bell, D. J., & Ruthven, I. (2004). Searcher's assessments of task complexity for web searching. In *European conference on information retrieval* (pp. 57–71).
- Beyer, H., & Holtzblatt, K. (1997). *Contextual design: defining customer-centered systems*. Elsevier.
- Beyer, H., & Holtzblatt, K. (1999). Contextual design. *Interactions*, 6(1), 32–42.
- Bic, L. (1985). Dataflow architectures for knowledge representation systems, afips. In *Proceedings of the national computer conference* (Vol. 54, pp. P503–P511).
- Bloomberg, L. D., & Volpe, M. (2018). *Completing your qualitative dissertation: A road map from beginning to end*. Sage Publications.
- Borgman, C. L. (1985). The user's mental model of an information retrieval system. In *Proceedings of the 8th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 268–273).
- Borlund, P. (2003). The concept of relevance in ir. *Journal of the American Society for information Science and Technology*, 54(10), 913–925.
- Borlund, P. (2010). The cognitive viewpoint: The essence of information retrieval interaction. *The Janus faced scholar. A festschrift in honor of Peter Ingwersen*, 23–34.

- Borlund, P., & Dreier, S. (2014). An investigation of the search behaviour associated with Ingwersen's three types of information needs. *Information Processing & Management*, 50(4), 493–507.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2), 77–101.
- Brookes, B. C. (1980). The foundations of information science. part i. philosophical aspects. *Information Scientist*, 2(3-4), 125–133.
- Buckley, J. W., Buckley, M. H., & Chiang, H.-F. (1976). *Research methodology and business decisions*. Canada: National Association of Accountants.
- Budanovic, M. P., & Žumer, M. (2015). Investigating mental models of cataloguers as the first step towards the development of intuitive cataloguer's tools. In *Re:inventing Information Science in the Networked Society. Proceedings of the 14th International Symposium on Information Science (ISI 2015), Zadar, Croatia, 19th–21st May*. (pp. 106–117).
- Byström, K. (1999). *Task complexity, information types and information sources: examination of relationships* (Unpublished doctoral dissertation). Acta Universitatis Tamperensis 688. University of Tampere, Department of Information Studies, Thesis for the degree of Doctor of Social Sciences.
- Byström, K., & Hansen, P. (2005). Conceptual framework for tasks in information studies. *Journal of the Association for Information Science and Technology*, 56(10), 1050–1061.
- Byström, K., & Järvelin, K. (1995). Task complexity affects information seeking and use. *Information processing & management*, 31(2), 191–213.
- Campbell, D. J. (1988). Task complexity: A review and analysis. *Academy of management review*, 13(1), 40–52.
- Carbonell, J. R. (1970). AI in CAI: An artificial-intelligence approach to computer-assisted instruction. *IEEE transactions on man-machine systems*, 11(4), 190–202.
- Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. Sage.
- Chen, P. P.-S. (1976). The entity-relationship model—toward a unified view of data. *ACM Transactions on Database Systems (TODS)*, 1(1), 9–36.
- Cleary, M., Horsfall, J., & Hayter, M. (2014). Data collection and sampling in qualitative research: does size matter? *Journal of Advanced Nursing*, 70(3), 473–475.
- Cleverdon, C. W. (1991). The significance of the Cranfield tests on index languages. In *Proceedings of the 14th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 3–12).

- Cole, C. (2011). A theory of information need for information retrieval that connects information to knowledge. *Journal of the Association for Information Science and Technology*, 62(7), 1216–1231.
- Cole, C. (2015). Information need and the beginning of information search. In *Encyclopedia of information science and technology, third edition* (pp. 4117–4128). IGI Global.
- Cole, C., Julien, C.-A., & Leide, J. E. (2010). An associative index model for the results list based on vannevar bush's selection concept. *Information Research: An International Electronic Journal*, 15(3), n3.
- Corbin, J. M., & Strauss, A. L. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative sociology*, 13(1), 3–21.
- Corbin, J. M., & Strauss, A. L. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory (3rd ed.)*. SAGE.
- Coyle, K. (2015). FRBR, twenty years on. *Cataloging & Classification Quarterly*, 53(3-4), 265–285.
- Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches (3rd ed.)*. SAGE.
- Dervin, B. (1981). Mass communicating: changing conceptions of the audience. In *Public communication campaigns* (p. 71-87). SAGE.
- Dervin, B. (1983a). Information as a user construct: The relevance of perceived information needs to synthesis and interpretation. In *Knowledge structure and use: Implications for synthesis and interpretation* (pp. 155–183). Temple University Press.
- Dervin, B. (1983b). *An overview of sense-making research: Concepts, methods, and results to date. Paper presented to the International Communication Association annual meeting, Dallas, Texas.*
- Dervin, B. (1992). From the mind's eye of the user: The sense-making qualitative-quantitative methodology. *Qualitative research in information management*, 9, 61–84.
- Dervin, B. (1998). Sense-making theory and practice: an overview of user interests in knowledge seeking and use. *Journal of knowledge management*, 2(2), 36–46.
- Dervin, B., & Dewdney, P. (1986). Neutral questioning: A new approach to the reference interview. *Research Quarterly*, 25(4), 506–513.
- Dervin, B., & Nilan, M. (1986). Information needs and uses. *Annual review of information science and technology*, 21, 3–33.

- De Vries, A. P., Vercoustre, A.-M., Thom, J. A., Craswell, N., & Lalmas, M. (2007). Overview of the inex 2007 entity ranking track. In *International workshop of the initiative for the evaluation of xml retrieval* (pp. 245–251).
- Diriye, A., Wilson, M. L., Blandford, A., & Tombros, A. (2010). Revisiting exploratory search from the hci perspective. *HCIR 2010: Proceedings of the Fourth Workshop on Human-Computer Interaction and Information Retrieval*, 99.
- Doerr, M. (2005). The CIDOC CRM, an ontological approach to schema heterogeneity. In *Semantic interoperability and integration*.
- Doerr, M. (2009). Ontologies for cultural heritage. In *Handbook on ontologies* (pp. 463–486). Springer.
- Doerr, M., & Crofts, N. (1999). Electronic esperanto—The Role of the oo CIDOC Reference Model. *Proceedings of the ICHIM '99, Washington DC, 22–26 September 1999*.
- Doerr, M., Ore, C.-E., & Stead, S. (2007). The CIDOC conceptual reference model: a new standard for knowledge sharing. In *Tutorials, posters, panels and industrial contributions at the 26th international conference on Conceptual modeling-Vol. 83* (pp. 51–56).
- Duff, W., & Johnson, C. (2001). A virtual expression of need: An analysis of e-mail reference questions. *The American Archivist*, 64(1), 43–60.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis*. MIT press Cambridge, MA.
- Farrokhnia, M., & Aalberg, T. (2016). Finding user need patterns in the world of complex semantic cultural heritage data. In *Research conference on metadata and semantics research* (pp. 187–192).
- Giannoulakis, S., Tsapatsoulis, N., & Grammalidis, N. (2018). Metadata for intangible cultural heritage. *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP 2018)*, 634-645.
- Gill, T. (2004). Building semantic bridges between museums, libraries and archives: The cidoc conceptual reference model. *First Monday*, 9(5).
- Grogan, D. J. (1991). *Practical reference work*. Library Assn Pub Limited.
- Gross, M. (1995). The imposed query. *RQ*, 35(2), 236–243.
- Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2), 199–220.
- Hajmoosaei, A., & Skoric, P. (2016). Museum ontology-based metadata. In *Semantic Computing (ICSC), 2016 IEEE Tenth International Conference on* (pp. 100–103).

- He, D., Brusilovsky, P., Ahn, J., Grady, J., Farzan, R., Peng, Y., ... Rogati, M. (2008). An evaluation of adaptive filtering in the context of realistic task-based information exploration. *Information Processing & Management*, 44(2), 511–533.
- Hennicke, S. (2017). *What is the Real Question? An Empirical-Ontological Approach to the Interpretative Analysis of Archival Reference Questions* (Unpublished doctoral dissertation). Humboldt-Universität zu Berlin, Philosophische Fakultät, Dissertation for the degree of Philosophiae Doctor (Ph.D.).
- Hjørland, B. (1997). *Information seeking and subject representation: An activity-theoretical approach to information science*. Greenwood Press.
- Hjørland, B. (2013). Theories of knowledge organization—theories of knowledge. *Knowledge organization*, 40(3), 169–181.
- Hofmann, K., Rijke, M., Huurnink, B., & Meij, E. (2009). A semantic perspective on query log analysis. In *In Working notes for the CLEF 2009 Workshop, Cortu*.
- Holmen, J., Ore, C.-E., & Eide, O. (2004). Documenting two histories at once: digging into archaeology. *BAR International Series*, 1227, 221–224.
- Holtzblatt, K., & Beyer, H. (2014). Contextual design: evolved. *Synthesis Lectures on Human-Centered Informatics*, 7(4), 1–91.
- IFLA study group on the Functional Requirements for Bibliographic Records. (1998). *Functional requirements for bibliographic records, final report*.
- Ingwersen, P. (1982). Search procedures in the library—analysed from the cognitive point of view. *Journal of documentation*, 38(3), 165–191.
- Ingwersen, P. (1986). Cognitive analysis and the role of the intermediary in information retrieval. *Intelligent information systems*, 206–237.
- Ingwersen, P. (1992). *Information retrieval interaction* (Vol. 246). Taylor Graham London.
- Ingwersen, P. (1994). Polyrepresentation of information needs and semantic entities elements of a cognitive theory for information retrieval interaction. In *SIGIR'94* (pp. 101–110).
- Ingwersen, P. (1996). Cognitive perspectives of information retrieval interaction: elements of a cognitive ir theory. *Journal of documentation*, 52(1), 3–50.
- Ingwersen, P., & Järvelin, K. (2005). *The turn: Integration of information seeking and retrieval in context*. Springer.
- Jacobs, R. L. (1985). *Naturalistic inquiry and qualitative methods: Implications for training and development. training & development research report*. Ohio State University.

- Jahoda, G. (1977). *The process of answering reference questions. a test of a descriptive model*. US Dept. of Health, Education and Welfare.
- Jamshed, S. (2014). Qualitative research method-interviewing and observation. *Journal of basic and clinical pharmacy*, 5(4), 87.
- Järvelin, K., & Repo, A. (1983). On the impacts of modern information technology on information needs and seeking: A framework. *Representation and exchange of knowledge as a basis of information processes*, 207–230.
- Kakali, C., Lourdi, I., Stasinopoulou, T., Bountouri, L., Papatheodorou, C., Doerr, M., & Gergatsoulis, M. (2007). Integrating dublin core metadata for cultural heritage collections using ontologies. In *International conference on dublin core and metadata applications* (pp. 128–139).
- Katz, L. S. (2013). *Electronic resources: use and user behavior*. Routledge.
- Kelly, G. A. (1969). *Clinical psychology and personality: The selected papers. edited by brendan maher*. Wiley.
- Kelly, G. A. (1970). A brief introduction to personal construct theory. *Perspectives in personal construct theory*, 1–29.
- Kelly, G. A. (2003). A brief introduction to personal construct theory. In *International handbook of personal construct psychology* (pp. 3–20). John Wiley & Sons.
- Kim, J. (2009). Describing and predicting information-seeking behavior on the web. *Journal of the Association for Information Science and Technology*, 60(4), 679–693.
- Kuhlthau, C. C. (1988). Developing a model of the library search process: Cognitive and affective aspects. *Rq*, 232–242.
- Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. *Journal of the American society for information science*, 42(5), 361.
- Kuhlthau, C. C. (1993a). A principle of uncertainty for information seeking. *Journal of documentation*, 49(4), 339–355.
- Kuhlthau, C. C. (1993b). *Seeking meaning*. Norwood, NJ: Ablex.
- Kvale, S., & Brinkmann, S. (2009). *Interviews: Learning the craft of qualitative research interviewing (2nd ed.)*. SAGE.
- Le Boeuf, P. (2003). *Brave new FRBR world*. Paper presented at the IFLA Meeting of Experts for an International Cataloguing Code, Frankfurt, 2003.

- Le Boeuf, P. (2006). *Using an ontology-driven system to integrate museum information and library information*. Paper presented at the Symposium on Digital Semantic Content across Cultures, Paris, the Louvre, 4-5 May.
- Le Boeuf, P. (2013). *Transforming FRBR into FRBRoo*. Lecture presented at Università La Sapienza, Rome, 10 May.
- Le Boeuf, P., Doerr, M., Ore, C., & Stead, S. (2018). Definition of the CIDOC Conceptual Reference Model, v. 6.2.3. *ICOM/CIDOC CRM SIG*.
- Lee, J. H., Renear, A., & Smith, L. C. (2006). Known-item search: Variations on a concept. *Proceedings of the Association for Information Science and Technology*, 43(1), 1–17.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry* (Vol. 75). Sage.
- Lourdi, I., & Papatheodorou, C. (2006). *Metadata policies for the description of digital folklore collections*. Paper presented at the Third International Conference of Museology & Annual Conference of AVICOM Mytilene, June 5 – 9, 2006.
- Marchionini, G. (1990). Effects of search and subject expertise on information seeking in a hypertext environment. In *Proceedings of the 53rd annual meeting of the asis, 1990*.
- Marchionini, G. (1995). *Information seeking in electronic environments*. Cambridge University Press, NY.
- Marchionini, G. (2006). Exploratory search: from finding to understanding. *Communications of the ACM*, 49(4), 41–46.
- Marchionini, G., Dwiggins, S., Katz, A., & Lin, X. (1993). Information seeking in full-text end-user-oriented search systems: The roles of domain and search expertise. *Library and information science research*, 15(1), 35–69.
- Markey, K. (1981). Levels of question formulation in negotiation of information need during the online presearch interview: A proposed model. *Information Processing & Management*, 17(5), 215–225.
- Merčun, T. (2012). *Visualization of results and navigation support in user interfaces of bibliographic information systems* (Unpublished doctoral dissertation). University of Ljubljana, Faculty of Arts, Department of Library and Information Science and Book Studies, Doctoral dissertation.
- Merriam, S. B., & Tisdell, E. J. (2015). *Qualitative research: A guide to design and implementation*. John Wiley & Sons.
- Miles, M. B., Huberman, A. M., & Saldana, J. (2014). *Qualitative data analysis*. Sage.

- Morris, R. C. (1994). Toward a user-centered information service. *Journal of the American Society for Information Science (1986-1998)*, 45(1), 20.
- Nicolaisen, J. (2009). Compromised need and the label effect: An examination of claims and evidence. *Journal of the American Society for Information Science and Technology*, 60(10), 2004–2009.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological review*, 84(3), 231.
- Niu, Y., & Winter, S. (2006). The effect of task type and information format on web searching performance. *ICIS 2006 Proceedings*, 52.
- Noc, M., & Žumer, M. (2014). Eliciting mental models of music resources: a research agenda. *Information Research*, 19(4).
- Norman, D. A. (2014). Some observations on mental models. In *Mental models* (pp. 7–14). Psychology Press.
- Noy, N. F., McGuinness, D. L., et al. (2001). *Ontology development 101: A guide to creating your first ontology*. Stanford knowledge systems laboratory technical report KSL-01-05 and Stanford medical informatics technical report SMI-2001-0880, Stanford, CA.
- Nunes, N. J. (2001). *Object modeling for user-centered development and user interface design: the wisdom approach* (Unpublished doctoral dissertation). University of Madeira, Thesis for the Degree of Doctor in Systems Engineering, specialty of Computer Science.
- Obrst, L. (2003). Ontologies for semantically interoperable systems. In *Proceedings of the twelfth international conference on information and knowledge management* (pp. 366–369).
- Oldman, D. (2014). The CIDOC Conceptual Reference Model (CIDOC-CRM): PRIMER. *CIDOC-CRM official web site*.
- Paisley, W. J., & Parker, E. B. (1965). Information retrieval as a receiver-controlled communication system. In *Proceedings of the symposium on education for information science* (pp. 23–31).
- Patton, M. Q. (2015). *Qualitative research and methods: Integrating theory and practice*. London: Sage Publications Ltd.
- Pettigrew, K. E. (2000). Lay information provision in community settings: How community health nurses disseminate human services information to the elderly. *The Library Quarterly*, 70(1), 47–85.
- Pharo, N. (2002). *The sst method schema: a tool for analysing work task-based web information search processes*. Tampere University Press.

- Pharo, N. (2004). A new model of information behaviour based on the search situation transition schema. *Information Research: An International Electronic Journal*, 10(1), n1.
- Pirolli, P. (2007). *Information foraging theory: Adaptive interaction with information*. Oxford University Press.
- Pisanski, J., & Žumer, M. (2010a). Mental models of the bibliographic universe. part 1: mental models of descriptions. *Journal of Documentation*, 66(5), 643–667.
- Pisanski, J., & Žumer, M. (2010b). Mental models of the bibliographic universe. part 2: comparison task and conclusions. *Journal of Documentation*, 66(5), 668–680.
- Quillan, M. R. (1966). *Semantic memory* (Tech. Rep.). BOLT BERANEK AND NEWMAN INC CAMBRIDGE MA.
- Raven, M. E., & Flanders, A. (1996). Using contextual inquiry to learn about your audiences. *ACM SIGDOC Asterisk Journal of Computer Documentation*, 20(1), 1–13.
- Ritchie, J., & Lewis, J. (Eds.). (2003). *Qualitative research practice: A guide for social science students and researchers*. sage.
- Riva, P., Le Boeuf, P., & Žumer, M. (2017). *IFLA Library Reference Model*. IFLA.
- Ruthven, I., & Kelly, D. (2011). *Interactive information seeking, behaviour and retrieval*. Facet Publishing.
- Saldaña, J. (2014). *Thinking qualitatively: Methods of mind*. Sage Publications.
- Saldaña, J. (2015). *The coding manual for qualitative researchers*. Sage.
- Sami, L. K., & Uplaonkar, S. S. (2015). Information seeking behavior among horticultural scientists in dr. ysr horticultural university, venkatramangudem. *Asian Journal of Information Science and Technology*, 5(1), 11–14.
- Saracevic, T. (1997). The stratified model of information retrieval interaction: Extension and applications. In *Proceedings of the annual meeting-american society for information science* (Vol. 34, pp. 313–327).
- Savolainen, R. (1993). The sense-making theory: Reviewing the interests of a user-centered approach to information seeking and use. *Information processing & management*, 29(1), 13–28.
- Schacter, J., Chung, G. K., & Dorr, A. (1998). Children's internet searching on complex problems: performance and process analyses. *Journal of the Association for Information Science and Technology*, 49(9), 840–849.
- Schamber, L., Eisenberg, M. B., & Nilan, M. S. (1990). A re-examination of relevance: toward a dynamic, situational definition\*. *Information processing & management*, 26(6), 755–776.

- Schutz, A. (1970). *Reflections on the problem of relevance*. Yale University Press.
- Shannon, C., & Weaver, W. (1949). *The mathematical theory of communication*. University of Illinois press, Urbana.
- Sowa, J. F. (1991). *Principles of semantic networks: Exploration in the representation of knowledge*, Jf sowa and a. borgida, eds. Morgan Kaufmann, Jan.
- Sowa, J. F. (2000). *Knowledge representation: logical, philosophical, and computational foundations*. Brooks/Cole Pacific Grove.
- Spencer, L., & Ritchie, J. (2002). Qualitative data analysis for applied policy research. In *Analyzing qualitative data* (pp. 187–208). Routledge.
- Stasinopoulou, T., Bountouri, L., Kakali, C., Lourdi, I., Papatheodorou, C., Doerr, M., & Gergatsoulis, M. (2007). Ontology-based metadata integration in the cultural heritage domain. In *International conference on asian digital libraries* (pp. 165–175).
- Stead, S., Doerr, M., & Theodoridou, M. (2015). *CRMinf: the argumentation model*.
- Swanson, D. R. (1986). Subjective versus objective relevance in bibliographic retrieval systems. *The Library Quarterly*, 56(4), 389–398.
- Takahashi, J., Kushida, T., Hong, J.-K., Sugita, S., Kurita, Y., Rieger, R., . . . Loverance, R. (1998). Global digital museum: multimedia information access and creation on the internet. In *Proceedings of the third acm conference on digital libraries* (pp. 244–253).
- Tallerås, K., Dahl, J. H. B., & Pharo, N. (2018). User conceptualizations of derivative relationships in the bibliographic universe. *Journal of Documentation*, 74(4), 894–916.
- Taylor, R. S. (1962). The process of asking questions. *Journal of the Association for Information Science and Technology*, 13(4), 391–396.
- Taylor, R. S. (1967). *Question-negotiation an information-seeking in libraries*. (Tech. Rep.). LEHIGH UNIV BETHLEHEM PA CENTER FOR INFORMATION SCIENCE.
- Taylor, R. S. (1968). Question-negotiation and information seeking in libraries. *College and Research Libraries*, 29(3), 178–194.
- Taylor, R. S., & Taylor, R. S. (1986). *Value-added processes in information systems*. Greenwood Publishing Group.
- Toms, E. G. (2011). Task-based information searching and retrieval. *Interactive Information Seeking, Behaviour and Retrieval*, 43–59.
- Toms, E. G. (2015). Complex tools for complex tasks. In *Proceedings of the first international workshop on supporting complex search tasks, volume 1338 of ceur workshop proceedings*. ceur-ws.org.

- Vakkari, P. (2001). A theory of the task-based information retrieval process: a summary and generalisation of a longitudinal study. *Journal of documentation*, 57(1), 44–60.
- Vakkari, P. (2010). Exploratory searching as conceptual exploration. *Proceedings of HCIR*, 24–27.
- Vlahavas, I., & Bassiliades, N. (1998). *Parallel, object-oriented, and active knowledge base systems* (Vol. 11). Springer Science & Business Media.
- Voigt, M. J. (1961). Scientists' approaches to information. *ACRL Monograph*(24).
- Wang, P. (1999). Methodologies and methods for user behavioral research. *Annual review of information science and technology (ARIST)*, 34, 53–99.
- Weinstein, P. C. (1998). Ontology-based metadata. In *Proceedings of the third acm conference on digital libraries, june 23 - 26, 1998* (pp. 254–263). ACM.
- Wersig, G. (1971). *Information-kommunikation-dokumentation: Ein beitrag zur orientierung der informations-u. dokumentationswissenschaften*. Verlag Dokumentation.
- White, R. W., & Roth, R. A. (2009). Exploratory search: Beyond the query-response paradigm. *Synthesis lectures on information concepts, retrieval, and services*, 1(1), 1–98.
- Wildemuth, B. M. (2016). *Applications of social research methods to questions in information and library science*. ABC-CLIO.
- Wildemuth, B. M., & Freund, L. (2012). Assigning search tasks designed to elicit exploratory search behaviors. In *Proceedings of the symposium on human-computer interaction and information retrieval* (p. 4).
- Wildemuth, B. M., & O'Neill, A. L. (1995). The “known” in known-item searches: Empirical support for user-centered design (research note). *College & Research Libraries*, 56(3), 265–281.
- Wilson, P. (1973). Situational relevance. *Information storage and retrieval*, 9(8), 457–471.
- Wilson, T. D. (1984). The cognitive approach to information-seeking behaviour and information use. *Social Science Information Studies*, 4(2-3), 197–204.
- Wilson, T. D. (1999). Models in information behaviour research. *Journal of documentation*, 55(3), 249–270.
- Žumer, M., & Le Bœuf, P. (2006). Conceptual models: museums & libraries: towards an object-oriented formulation of FRBR aligned on the CIDOC CRM ontology. In *New tools and new library practices, 30th ELAG seminar (Bucharest 2006)*.
- Žumer, M., & Riva, P. (2017). IFLA LRM-finally here. In *International conference on dublin core and metadata applications* (pp. 13–23).

- Žumer, M., Salaba, A., & Zhang, Y. (2012). User Verification of the FRBR Conceptual Model and Testing of FRBR Prototypes. In *International conference on asian digital libraries* (pp. 148–151).



Appendix A

## **Interview guide**

## **Interview Guide**

I am interested in what the information needs of users of libraries, archives and museums are; how they understand the content in the information systems they use; and how they interpret the search results. I realize that these information needs and preferences differ from user to user based on multiple factors such as professional background, research interests or search tasks. It might also vary from information item to information item. I would like to understand these differences and find the common patterns in order to help to have a better information representation that satisfies different needs of users in different domains (libraries, archives and museums).

### **Brief introductory questions**

- Now that you briefly understand the study, I would like to start by asking you to tell me a bit about yourself (such as age, work or professional background and the length of your research experience).
- What are your research interests?
- What is your current specific research area? And, how long have you been working on it?
- Which information systems and what kinds of information resources do you usually use for your researches?

### **The recent search experience**

- Would you please think about your recent information need and tell me about the related search you have done in an information system or internet?
- What query (queries) did you use and why? What did you expect to find through this query?
- How much did the result of your search satisfy you? How relevant was it?
- What did you pick up from the result list? And why?
- What else did you expect to see in the result? (regardless of the kind of information system and its limitations) What should not be included in the result?
- How would you prioritize the result? How would you prefer the result being organized (represented) and for what purposes?
- What did you find at the end? What led you to reformulate your query?(if applicable)
- How would you explore this topic and based on what features you would do it?
- Would you like to add anything more?

Appendix B

# **Consent Form**

# **Request for participation in research project**

## **Background and Purpose**

As part of my PhD project in Library and Information Science at Oslo and Akershus University College of Applied Sciences, I am going to do a qualitative research on users' needs and preferences. The purpose of this research is to understand users' needs and preferences in different domains of memory institutions (Libraries, Archives and Museums).

For this research, a sample of PhD students from different disciplines considering the potential users of archives, libraries and museums has been selected. You are being kindly invited to take part in this research because you are one of the potential participants.

## **What does participation in the project imply?**

If you agree to participate in my research, I will conduct an interview with you at a time and location of your choice. The interview will involve questions about the kinds of information you expect and prefer to see when you search in an information system.

I expect the interview will last about 60-90 minutes. With your permission, I will audio record and take notes during the interview. The recording is to accurately record the information you provide, and will be used for transcription purposes only. No identifying information will be recorded on the transcription. If you do grant permission for this conversation to be recorded, you have the right to revoke recording permission and/or end the interview at any time.

## **What will happen to the information about you?**

The information recorded is confidential, and no one else except me will have access to the information documented during your interview. If the results of this study are published or presented, individual names and other personally identifiable information will not be used. Therefore, participants will not be recognizable in the publications.

This project is scheduled for completion by 1 November 2016. All the interview recordings will be destroyed one year after that date. Until they are destroyed, they will be stored in a secure workspace and in a password protected computer file to which only the researcher (me) have access.

## **Voluntary participation**

It is voluntary to participate in the project, and you can at any time choose to withdraw your consent without stating any reason. If you decide to withdraw, all your personal data will be made anonymous.

If you have any questions concerning the project, please contact Maliheh Farrokhnia (principal researcher) at Phone number 0047 95905408. You can also contact my main supervisor Professor Nils Pharo at 0047 67238302 if you have any further queries.

The study has been notified to the Data Protection Official for Research, Norwegian Social Science Data Services.

## **Consent for participation in the study**

I have received information about the project and am willing to participate

-----

(Signed by participant, date)



Appendix C

# **Approval from Norwegian Center for Research Data, NSD**



Harald Hårfagres gate 29  
N-5007 Bergen  
Norway  
Tel: +47-55 58 21 17  
Fax: +47-55 58 96 50  
nsd@nsd.uib.no  
www.nsd.uib.no  
Org.nr. 985 321 884

Maliheh Farrokhnia

Institutt for arkiv, bibliotek- og informasjonsfag Høgskolen i Oslo og Akershus

Pilestredet 48

0167 OSLO

Vår dato: 16.04.2015

Vår ref: 42603 / 3 / KH

Deres dato:

Deres ref:

## TILBAKEMELDING PÅ MELDING OM BEHANDLING AV PERSONOPPLYSNINGER

Vi viser til melding om behandling av personopplysninger, mottatt 04.03.2015. Meldingen gjelder prosjektet:

42603

*A framework for interacting with FRBRoo/ CIDOC CRM data: With emphasis on heterogeneous needs*

*Behandlingsansvarlig*

*Høgskolen i Oslo og Akershus, ved institusjonens øverste leder*

*Daglig ansvarlig*

*Maliheh Farrokhnia*

Etter gjennomgang av opplysninger gitt i meldeskjemaet og øvrig dokumentasjon, finner vi at prosjektet ikke medfører meldeplikt eller konsesjonsplikt etter personopplysningslovens §§ 31 og 33.

Dersom prosjektopplegget endres i forhold til de opplysninger som ligger til grunn for vår vurdering, skal prosjektet meldes på nytt. Endringsmeldinger gis via et eget skjema, <http://www.nsd.uib.no/personvern/meldeplikt/skjema.html>.

Vedlagt følger vår begrunnelse for hvorfor prosjektet ikke er meldepliktig.

Vennlig hilsen

Vigdis Namtvedt Kvalheim

Kjersti Haugstvedt

Kontaktperson: Kjersti Haugstvedt tlf: 55 58 29 53

Vedlegg: Prosjektvurdering

*Dokumentet er elektronisk produsert og godkjent ved NSDs rutiner for elektronisk godkjenning.*

*Avdelingskontorer / District Offices:*

*OSLO:* NSD, Universitetet i Oslo, Postboks 1055 Blindern, 0316 Oslo. Tel: +47-22 85 52 11. nsd@uio.no

*TRONDHEIM:* NSD, Norges teknisk-naturvitenskapelige universitet, 7491 Trondheim. Tel: +47-73 59 19 07. kyrre.svarva@svt.ntnu.no

*TROMSØ:* NSD, SVF, Universitetet i Tromsø, 9037 Tromsø. Tel: +47-77 64 43 36. nsdmaa@sv.uit.no

# Personvernombudet for forskning



## Prosjektvurdering - Kommentar

---

Prosjektnr: 42603

Based on the information we have received about the project, the Data Protection Official cannot see that the project will entail a processing of personal data by electronic means, or an establishment of a manual personal data filing system containing sensitive data. The project will therefore not be subject to notification according to the Personal Data Act.

The Data Protection Official presupposes that all information processed using electronic equipment in the project is anonymous.

Anonymous information is defined as information that cannot identify individuals in the data set in any of the following ways:

- directly, through uniquely identifiable characteristic (such as name, social security number, email address, etc.)
  
- indirectly, through a combination of background variables (such as educational institution, gender, age, etc.)
- through a list of names referring to an encryption formula or code, or
- through identifying information on sound recordings.