A Formalization of the »Digital Methods« – Supporting Comprehensible Access to the Novel Web Science Research Field

submitted to obtain the degree of

Master of Science (M.Sc.)

at

Cologne University of Applied Sciences

Institute of Informatics

Course of Studies: Web Science

by

Miriam Schmitz (11053034)

Subbelrather Str. 351

50825 Köln

miry.schmitz@gmail.com

Cologne, April 2014

First Supervisor: Prof. Dr. Kristian Fischer
(Cologne University of Applied Sciences)

Second Supervisor: Dr. Wolter Pieters
(Delft University of Technology)

Abstract

This paper is grounded in the emerging field of web science and shall contribute to its further classification and demarcation by illustrating the current state of »web-native research methods«. It builds upon an initial arraying work of Richard Rogers, who coined the term »Digital Methods« for research with methods that were »born« in the web, and illustrated and organized them in his eponymous book in 2013. This paper attempts to develop a more appropriate illustration of the Digital Methods by following the web's very own, hypertextual, network-like nature, in particular by construing an ontological representation on the base of the Web Ontology Language (OWL). By virtue of decomposing the book into granular information units and their subsequent reassembly into OWL entities, immediate access to the entire knowledge domain can be provided, and coherencies, interrelations and distinctions between concepts become apparent. The ontology's structure was induced narrowly along the provided examples of research projects and subsequently clustered in topic groups, of which the three most important ones were (a) the Digital Methods as an arraying space of web-native methodology, (b) a collection of concrete applications of these Digital Methods in research projects, and (c) a hierarchical scheme of traditional sciences with a distinct interest in answering research questions with help of Digital Methods. Subsequently, the ontology was evaluated in three general dimensions: Deriving user stories and scenarios provided means to validate the utilization quality; the accuracy and reliability of the resulting structure was validated with help of a control group of web-native research projects; and process control instruments served as a validator for the ontology's correctness. Despite the ontology itself, this paper also resulted in a first interpretation of the produced information: Statements about research practise in social science, politics and philosophy were as possible as findings about commonly applied varieties of methods. Concluding, the present paper proposes a process of ontology engineering, an evaluation of the ontology's value, and an interpretation of the ontology's content.

This work was conducted using the Protégé resource, which is supported by grant GM10331601 from the National Institute of General Medical Sciences of the United States National Institutes of Health.

Table of Content

1 Introduction 7
1.1 Research Problem 7
1.2 Motivation 12
1.3 Method 16
1.4 Structure of this Document 19

2 Identifying Essential Use 21
2.1 Essential Use Cases 21
2.2 Stakeholder Analysis 22
2.3 User Stories 24

3 Epistemological and Methodological Foundations 26
3.1 Introduction 26
3.2 Epistemological Foundations – Definitions & Differentiation of Domain 27
3.3 Methodological Foundations 30
3.4 Ontological Suppositions 34

4 Implementation 38
4.1 Approach to Induction 38
4.2 Corresponding OWL Concepts 40
4.3 Exemplary Initial Collection 46
4.4 Additional Problems Solved 49

5 Results 51
5.1 The General Structure of the Digital Methods Domain 51
5.2 Prospective Use 56

6 Evaluation 57
6.1 Introduction 57
6.2 Result Validity 58
6.3 Process Reliability 64
6.4 Utilization Quality 68
6.5 Conclusion of Evaluation 74

7 Interpretation 76
7.1 Introduction 76
7.2 The Current State of the Digital Methods Research Field 77
7.3 Conclusion of Interpretation 85

8 Discussion & Conclusion 86
8.1 Conclusion 86
8.2 Outlook 89

9 Resources 91
Declaration in Lieu of Oath 95

Illustrations Index

Illustration 1-1: The Fields of Investigation (own illustration) ... 16
Illustration 1-2: Simplified Reversed Tree Structure of the Digital Methods Ontology (own illustration) ... 18
Illustration 2-1: Essential Use Cases (own illustration) ... 21
Illustration 2-2: General Framework for Studying Ontological Mediation (Anticoli & Toppano 2013: 25) ... 22
Illustration 2-3: User-focused Evaluation Process from Essential Use Cases To User Stories to Scenarios (own illustration) ... 24
Illustration 3-1: Anthropological Scientific Disciplines Demarcation for English and German Language Spaces, According to Wikipedia (own illustration) ... 33
Illustration 4-1: The Triangle of Digital Methods, Research Questions, and Applications in Studies (own illustration) ... 39
Illustration 4-2: Necessary Relationships (Properties) of Superclasses (own illustration) ... 45
Illustration 4-3: Necessary Relationships (Properties) of Superclasses Extended (own illustration) ... 46
Illustration 4-4: Ontology Tree Structure of Collection Process (own illustration) ... 48
Illustration 5-1: The Digital Methods Book After Inspection (photography) ... 51
Illustration 5-2: Map of all Ontology Items (Protégé export) ... 52
Illustration 5-3: Neighbourhood of the *ResearchProject* class and all contained individuals (Protégé export) ... 53
Illustration 5-4: The *DigitalMethods* Class Thread with all Contained Entities (Protégé export) ... 55
Illustration 5-5: The *ResearchDomain* Class Thread with all Contained Entities (Protégé export) ... 56
Illustration 6-1: Domains and Ranges of Toplevel Classes (own illustration) ... 59
Illustration 6-2: False Attributions of Individuals to Class ResearchDomain (Protégé screenshot) ... 60
Illustration 6-3: Reasoner Explanation View (Protégé screenshot) ... 61
Illustration 6-4: Individual Object Property Assertions View (Protégé screenshot) ... 61
Illustration 6-5: Exemplary User Journey for User Story 1 – Desired Scenario (Protégé export) ... 70
Illustration 6-6: Exemplary User Journey Around »Political Geography Online« for User Story 2 – Desired Scenario (Protégé export) ... 70
Illustration 6-7: Exemplary User Journey for User Story 3 – Desired Scenario (Protégé export) ... 73
Illustration 6-8: Exemplary Extract of the Ontology's XML Export as Potentially Used in User Story 4 – Desired Scenario (Protégé export) ... 74
Illustration 7-1: Subclasses of the Domain of Social Research (Protégé screenshot) ... 79
Illustration 7-2: Philosophy Class with two Individuals (Protégé screenshot) ... 81
Illustration 7-3: Subclasses of the Politics Class (Protégé screenshot) ... 81
Illustration 7-4: Individuals of the Politics Class (Protégé screenshot) ... 82
Illustration 7-5: Subclasses and Individuals in the DigitalMethods Superclass (Protégé screenshot) ... 83
Illustration 7-6: Number of Projects Conducted During Course of Years 1999 - 2011 (own illustration) ... 84

Tables Index

Table 2-1: Stakeholders with interest in the Digital Methods Ontology 23

Table 2-2: User Story 1 – Retrieve Methods Suitable for Reuse 25

Table 2-3: User Story 2 – Integrate Own Findings in Appropriate Location 25

Table 2-4: User Story 3 – Retrieve Information about own Project and Evaluate Correctness 25

Table 2-5: User Story 4 – Comprehend Knowledge Domain and Reuse it for Broader Context 25

Table 6-1: Control Group Object 1 – #Ausvotes: Twitter Activity Patterns Across Electorates 65

Table 6-2: Control Group Object 2 – Social Media as a Measurement Tool of Depression in Populations 65

Table 6-3: Control Group Object 3 – Traditional Media Seen from Social Media 66

Table 6-4: Control Group Object 4 – The Geographically Uneven Coverage of Wikipedia 66

Table 6-5: Control Group Object 5 – Top 10 Twitter Languages in London 67

Table 6-6: Scenarios for User Story 1 – Retrieve Methods Suitable for Reuse 69

Table 6-7: Scenarios for User Story 2 – Integrate Own Findings in Appropriate Location 71

Table 6-8: Scenarios for User Story 3 – Retrieve Information about own Project and Evaluate Correctness 72

Table 6-9: Scenarios for User Story 4 – Comprehend Knowledge Domain and Reuse it for Broader Context 73

1 Introduction

»[The World Wide Web] has spread inexorably into other scientific disciplines, academe in general, commerce, entertainment, politics and almost anywhere where communication serves a purpose«
(Berners-Lee et al. 2006b: 2).

1.1 Research Problem

It is evident and undisputed that the web has arrived in every condition of life, whether it is in social interaction, individual and mass communication, in epistemological and ethnological questions or in labour environments and economical strategies. In parallel, many scientific studies have been conducted that are designated as web-based or web-focused. Two general perspectives may be identified here, as shown by Scherfer & Volpers:

1) Studies that evaluate the web as a *medium*, as a room of social interaction and human behaviour;
2) Studies that investigate the *technical structure* of the web and identify possibilities of improvements (2013: 11).

Both perspectives evolved somehow as subcategories of established sciences, providing approved methods and tools to apply when researching web phenomena: When evaluating the web as a social interaction space, methods from social science and its various branches are applied; it is for example common to transfer *Content Analysis*, a method from communication science, into research with web content. It is also natural to make use of methods known from traditional, computer science based *Network Theory* to investigate links in the world wide web. The example of So*cial Network Analysis* demonstrates the close affiliation of both dimensions; they both provide interconnected insights on human and technical levels, and each may make use of the counterpart's methodology or contribute to the counterpart's body of knowledge.

This paper promotes the establishment and independent investigation of a third perspective, which is currently perceived a subset of perspective one (ibid.): Studies that use the web as a *source* of perceptions about society and culture. The difference to the medium-driven or technical-focused perspectives one and two is occasionally marginal, yet important: the third dimension consists only activities of social research *with* the web, utilizing methods *that would not exist* without the web. These methods are referred to by Richard Rogers as »Digital Methods«, and described as the following:

> »This book presents a methodological outlook for research with the web. As such it is a proposal to reorient the field of Internet-related research by studying and repurposing what I term the methods of the medium, or perhaps more straightforwardly methods embedded in online devices. (…) The purpose is (…) to think along with them, and learn how they handle hyperlinks, hits, likes, tags, datestamps, and other natively digital objects. By continually thinking along with the devices and the objects they handle, digital methods, as a research practice, strive to follow the evolving methods of the medium« (Rogers 2013: 1).

Following this description, Digital Methods are e.g. the investigation of Google search phrases per location to predict flu outbreaks (ibid.: 22), or the analysis of country-specific Google results to discover the most significant right types in those very countries (ibid.: 106). Other examples from the web science discipline – outside of Rogers and his Digital Methods Initiative[1] – are numerous. At the time writing this paper, the »Proceedings of the 5th Annual ACM Web Science Conference« of 2013 were published (ACM 2014), giving more current examples of studies that apply web-native methods: In their paper »Social Media as a Measurement Tool of Depression in Populations«, De Choudhury, Counts & Horvitz examine »the potential for leveraging social media postings as a new type of lens in understanding depression in populations« (2013: 47). In a multi-step method, they gathered a large data set of Twitter postings created by individuals diagnosed with depression, then developed a probabilistic model trained on this corpus, and finally built a social media depression index with indicators for geographical, demographic and seasonal patterns of depression. It could be found the data correlated strongly with depression statistics reported by the Centers for Disease Control and Prevention (CDC).

Another project by An et al. attempts to understand media supply and demand landscapes in order to develop effective marketing strategies. By analysing Twitter, where »users actively follow a wide set of media sources, form interpersonal networks, and propagate interesting stories to their peers« (2013: 1), media subscription and interaction patterns, which had previously been hidden behind media corporations' databases, become visible.

The key to those studies is not the nature of the search engine or the online social network itself, but rather the data that is produced by *using* it, and that can be utilized to answer questions about society online and offline.

The distinction between the previously introduced perspectives one and two (in which the web is a medium resp. a technical construct) on the one hand and three (in which the web is a source of data) on the other hand is hence rather a methodological than an epistemological paradigm: Whereas the perceptions gained through research may

[1] The Digital Methods Initiative (DMI) is a research collaboration of several Dutch institutes and »a New Media PhD (training) program as well as a New Media research group in Media Studies, University of

concern the same *domain* in all perspectives, the *methodological set* of perspective three is clearly limited to using web-native data and hence pretty well-defined – albeit solely when detached from the generalized »parent« perspective one or two. Insofar, the initial approach to having only two dimensions is legitimate, but not sufficient for a clear understanding. Furthermore, the initial bilateral division into general social science and computer science shows very well that the conductors of web-native research projects may come from multiple disciplines and have varying research intentions. Establishing a third perspective may contribute to a clearer distinction and definition of one uniting »roof discipline« of all research that is based upon Digital Methods. This roof discipline will reside in web science, an evolving discipline that seems tailor-made for the outlined research field, as already stressed by Gloria et al.:

> »As the Internet continues to provide both an object of study and research tools, it raises many questions for methodologies of Web Science research. We now live in an era where big data, abundant data, and accessible data exists and where relational information is its most relevant characteristic. (…) We must find new ways to identify, refine and contextualize data. For Web Science, mastering data to scale while grounding it in viable social theory remains disjointed« (Gloria et al. 2013).

Defining and illustrating Digital Methods is one approach to overcoming this disjointedness of data and social theories. Still, just as complicated as the differentiation of the Digital Methods research field appears, as difficult is its comprehensive illustration. What is a natively digital research method, and what not? What are commonalities of certain research projects based on web-native data, and where do they differ? What motivates a conductor to approach a research question by solely focusing on Digital Methods? An initial hypothesis of this paper is: It is likely that all studies to fall under this third perspective differ significantly from each other in terms of research intention and original scientific »roof« domain due to the scientific perspective of their respective multidisciplinary conductors, whereas they might have only one common denominator: the web as a basis for data assessment.

This would not be a problem in itself, but it complicates the access to knowledge in this field: If the research intention of the previous example of *Social Network Analysis* was to gain insights into the behaviour of individuals in online social networks, and the conductor of this study resided in the domain of sociology, how would a computer scientist find out about it? He himself may use the same method to gain insights into the evolving (technical) network structure of that very online social network, and to ground his own work upon. Obviously, there are numerous knowledge bases dedicated to certain disciplines, and the computer scientist may well retrieve the relevant study from the field of sociology in dedicated databases – on condition that he had either some cross-disciplinary expertise or a tangible idea of this one study or method. But the moment he wanted to randomly explore the field of *related* projects or methods, he

would be limited by the professional domain that surrounded the study, in this case social science. At best, he would explore all web-related social research, hereby possibly disregarding relevant items in other domains. A settlement of Digital Methods as an independent research field within the web science domain would support access to a thorough understanding. It is hence desirable to contribute to its establishment and perception by illustrating it in the completest possible way.

Rogers' book (Rogers 2013), presenting a considerably copious, annotated aggregation of web-based studies and methods, takes a step in this direction: Research projects are collected, described and sorted from a perspective of »the web«. Nevertheless, the book format in general is in a fundamental dilemma when describing any state of research: The author can only capture one »frozen« snapshot of a constantly transforming, ephemeral research state. By publishing a closed book, he therefore actively sets a caesura in the matter, indicating that »right now« was an appropriate moment to pause and retrospectively analyse the situation, or that right now was a finalized state of research. However, an emergent discipline is never in stagnation; its essentials, methods and applications are subjects to constant dialogue and transformation, and the caesura would be of a very artificial nature. In the concrete case of Rogers, further research within the Digital Methods domain during the course of time – by Rogers or any other research professional – will require changes of the book in order to

a) incorporate necessary additions to the incessantly infinite collection of projects or
b) allow for adjustments of the domain in case of evolving methodologies or methods.

A conceivable example to explain these requirements would be a future web-based technology with the same significance for research as the invention of Facebook or Twitter, of Smartphones or Smart TVs, that would e.g. allow for an unprecedented creation of online profiles, and that would make user data entirely accessible for research. New research projects would evolve in various scientific disciplines, and just as network theory was adapted for analysing Facebook, other traditional methods from suitable domains would be used to study this new phenomenon. By adapting and transforming these methods, a methodological change would be initiated, and scientific discourse would alter, resulting in the need to restructure the whole Digital Methods research field.

Thus, books suffer from a general inability to sufficiently represent an emergent discipline due to their inability for future scale[2].

Apart from publishing processes, another problem lies in perception: A book is analogue, and it is linear. Using the book format, Rogers presents his findings in a human logic, more precisely in a linear chain of reasoning with the motivation to convince a specific scientific community of new ideas and concepts. The value behind this is obvious in terms of scientific progress, yet it is debatable that the presentation form is the most appropriate for this very field: a linear text contains all relevant knowledge, but has to be discovered in a linear, possibly protracted intellectual process. This might hinder a thorough understanding and limit the depth of perception of the reader: Arguments are not grouped together by similarity, but follow a flow of argumentation; inferences are usually not drawn by the reader, but by the author. And ultimately: Knowledge is not composed according to user needs, but through the author's understanding of the best chronology.

A third problem with the book concerns, more concrete, the *content* of Digital Methods, which is simply ignored by the stiff, analogue book format. When debating hyperlink networks pointing to political situations (Rogers 2013: 6) as well as to archived states of the web (ibid.: 80), why not use a *hypertext* that shows this relationship? When discussing how online social network data is able to reveal new demographics, why not display this knowledge in a network graph? Despite these single studies, the possibilities of adjusting the output along the study subject apply to the meta level (the complete collection) as well: If there is one specific research field related to the web, but the scientific backgrounds are numerous, how can these relations to superior knowledge fields be displayed? Which representation would support a clear picture of the current state of this field, and allow for future integration of upcoming studies? If this knowledge area is about the web, why not allow for the web to »know« about it?

This paper attempts to develop a more appropriate representation of the web-native research field, and, as suggested in the previous section, it will learn from the web itself what to do: If any (not necessarily web-related) knowledge shall be represented – and the knowledge is in fact *about* the web, and about methods that were »born« here – then the most obvious solution that a web scientist could anticipate is a semantic web representation. More solutions were conceivable, but the charm of semantic technologies lies in the fact that to investigate the web, one makes use of the web's very own nature. Additionally, with using the Web Ontology Language (OWL), important conceptual challenges can be solved: The examples of discrepancy between

[2] Transformation in the matter of interest is usually solved by publishing new editions; however, this strategy is still not acknowledging external input by other researchers.

content and format ((a) the hyperlink, (b) the network, and (c) the relationships within the knowledge domain) prompt for

a) a graph illustration, where entities may have infinite relations to other entities,
b) a taxonomy, where classes have subclasses and superior classes and
c) machine-readable output.

OWL provides a solution for all three areas. With help of visualization tools based upon OWL, the graph illustration is provided; the language itself provides a well-elaborated concept of sorting knowledge into taxonomies, and is machine-readable by nature. As opposed to the linear reception that a book requires, reception of an ontology is associative. Hence, this paper attempts to build upon Rogers' findings and inductively constitute a Digital Methods ontology in the Web Ontology Language. By virtue of constituting an OWL ontology, it shall be possible to integrate all existing Digital Methods, the research projects in which they were applied, and their relations to traditional scientific domains. Due to the scalability of this data, it may prospectively be used for the further development of sorting web-related research or transformed for entirely different questions. By delivering the knowledge from the described book restrictions, it can also show comprehensively where the research interest of traditional science in the web accumulates and where there is a considerably weak coverage of using the web for research. As a side effect, the ontology built with OWL will output machine-readable data describing the research field of web-native methods.

The following research question, consolidating the illustrated perceptions, shall lead through the progress of this paper:

Is an ontological formalization appropriate to provide comprehensible access to the current state of the »Digital Methods« research field, and to visualize the connections among these very studies as well as their relations to established sciences?

1.2 Motivation

Besides the desired comprehensive illustration of a current state of research, a subordinate focus of this paper is put on the development of web science as a stand-alone scientific discipline. By illustrating the current status of one specific research field by means of a knowledge representation, this paper contributes to the meta-level investigation of the global web science field: Where are concentrations of investigation? Which questions or challenges have not yet been tackled sufficiently? What are the most important nodes to traditional sciences? The arising ontology will be grounded in the collection of Rogers' latest publication, »Digital Methods«, albeit solely on the collected research projects and respective research methods (as well as

occasional attempts of classifying or sorting them) rather than on the author's individual perception of trends, emerging research areas or general records of a web science evolvement. Nevertheless, this paper is grounded on the same overall philosophy: The perception of the web as more than an object of investigation[3] or a tool box for social, political or economical research. The web becomes the discipline, and its investigation may well be the primary aim of prospective research projects. Analogous, the Digital Methods are currently the objects of study, but may well prospectively be transferred into a discipline themselves: »Das umfassendere Ziel besteht darin, die Methoden der Internetforschung zu überarbeiten und damit einen neuen Studienzweig zu entwickeln: digitale Methoden« (Rogers 2011).

Although the intentions of the majority of studies may not underlay significant changes in the near future, the context in which they are grounded may transform from diverse traditional sciences with a special interest in the web to dedicated web professionals[4]. One may be disposed to accept this hypothetical transformation when looking at other examples of considerably young disciplines: Communications science, from which a lot of methodological input has been brought into web science, was itself a descendant of social sciences, and needed several years of scientific discourse to be established as a generic science. The German »Textwissenschaft« (discourse analysis) was established in the 70ies of the last century, among others by Teun A. van Dijk in his introductory work, where he disposes a liberation of the text from embeddedness in other sciences:

> »Wir erkennen daraus, daß des (sic!) 'Entstehen' einer neuen Wissenschaft für eine allgemeinere Analyse von Texten auf einer Linie mit Entwicklungen in mehreren Wissenschaftsdisziplinen liegt und damit die konsequente Fortführung einer Tendenz darstellt, Sprachgebrauch und Kommunikation interdisziplinär zu studieren« (van Dijk 1980: 1).

Independent from the development of discourse analysis, which was subsequently influenced by the rapid developments in information technologies, the author declares the general importance of new, independent disciplines for their ability to comprehend and explain current societal conditions and actions:

> »Wenn sich eine Wissenschaft von ihrer Mutterwissenschaft 'emanzipiert', dann liegt das nicht nur an den Fortschritten in den Untersuchungsmethoden oder den neuen Ergebnissen, sondern diese neue Wissenschaft stellt die Antwort dar auf bestimmte gesellschaftliche Entwicklungen« (ibid.: 2).

3 In the past, the web has for instance been investigated as a space of social interaction or as one of several mass-media phenomena.

4 Emerging scientific interest is usually manifested in university professions and programmes, such as the Web Science Master Programmes at Cologne University of Applied Sciences and Johannes Kepler University Linz, or the Research Group Data and Web Science at University of Mannheim, to name a few German-speaking projects.

Following this, a generic web science would not only be reasonable, but crucial: The information age would make it vital to develop an independent and tailored methodology and epistemology. The inventor of the world wide web, Tim Berners-Lee, adds to this an intrinsic motivation of sustaining a »healthy« web:

> »If we want to understand the architectural principles that have provided for its [the web's] growth; and if we want to be sure that it supports the basic social values of trustworthiness, privacy, and respect for social boundaries, then we must chart out a research agenda that targets the Web as a primary focus of attention« (Berners-Lee et al. 2006b: 1).

Some works from the recent past try to grasp this idea of Berners-Lee and establish a web science by gathering related methods[5], by providing distinct definitions[6] or a library of related research[7]. By virtue of constituting definitions and disassociations of certain study fields in numerous works, whereby strategy and methods differ significantly along with the research background, a general »shape« of web science is emerging. This paper shall contribute to a further differentiation with the bottom-up development of what may be a structuring, top-down-dispersed taxonomy afterwards. The high degree of interdisciplinarity, which a science of the web is subject of, will be taken into account by creating a network scheme of relationships of Digital Methods to traditional research domains. This will provide »anchors« to known concepts and established research, support familiarity of researchers with the matter, and by that ease the access to the novel research field.

As far as the individual – as opposed to the scientific community – is concerned, a reasonable, reliable future usage of the ontology as a research tool is desirable. Hence, this paper puts a primary focus on the improved representation of knowledge in the field of Digital Methods, which currently is only available through a printed book or electronically disposable equivalents; both analogue and digitized or natively digital[8] versions are comprised of linear[9] text that follows a human individual's logic and perception of coherencies as well as a one-dimensional narrative structure[10]. At current state, an all-embracing overview of the research field of Digital Methods is complicated not solely by the inability to »draw« direct lines between items (e.g. methods and all

---

[5] f.i. »Methoden der Webwissenschaft« by Scherfer & Volpers.
[6] f.i. »International Handbook of Internet Research« by Hunsinger, Klastrup & Allen.
[7] f.i. the »Digital Humanities« at University of Cologne.
[8] More on digitized and natively digital data can be found in Rogers (2013: 206-207); A commonly accepted difference is that digitized data is data transformed into a digital format, whereas natively digital, as its name implies, is data »born« in the digital – in this regard, in the web.
[9] Linear text as opposed to the multi-layered structure of hypertext and meta-text or text with non-linear and non-chronological references.
[10] According to van Dijk (1980: 150), every scientific discourse follows a certain argumentative superstructure; this theory was grounded in linguistics previous to the establishment of computer linguistics as a common sub field of computer science and refers to linear, human-readable and analogue output.

respective research projects), but also by the absence of navigation possibilities. Vannevar Bush, the originator of the antecedent hypertext, expresses similar concerns about regular text as far back as 1945, stating that research is significantly complicated by the mere inability to retrieve information out of text:

> »The prime action of use is selection, and here we are halting indeed. There may be millions of fine thoughts, and the account of the experience on which they are based, all encased within stone walls of acceptable architectural form; but if the scholar can get at only one a week by diligent search, his syntheses are not likely to keep up with the current Scene« (Bush 1997).

An alteration from linear text to »interactive« knowledge might be able to end the dilemma of differences in how the author versus the reader gives meaning to a domain, a phenomenon that some »constructivist« learning psychologists know as the following:

> »Construction of knowledge is the result of an active process of articulation and reflection within a context. […] Learning environments are constructivist only if they allow individuals or groups of individuals to make their own meaning for what they experience rather than requiring them to 'learn' the teacher's interpretation of that experience or content« (Jonassen et al. 1995).

Hypertextual perception enables an undirected, individually shaped navigation through information, and by that means allows for every reader to create his own story through resources. Transferred into the current paper, creating a hypertext or similar digital construct which enables some sort of navigation should facilitate the perception of the Digital Methods embedded in a greater scientific context. Additionally, OWL as a meta-language does not only provide a reasonable way of expressing this construct by formalization, it may even amplify the comprehension effect by the various possibilities of reuse. Examples of OWL applications for web services show a significant ease of perception of complex relationships in one domain; the *GoodRelations* Ontology for instance is utilized by O'Reilly among others to describe products and maintain disposability information in e-commerce websites (GoodRelations Wiki 2013); the *Music Ontology* »provides a model for publishing structured music-related data on your web site or through your API« (Pickering 2014), consisting of mainly business-directed meta-data about the music available on the web.

This research paper is not determined to provide an API to a Digital Methods Ontology nor will it fulfil the definition of W3C for a »good« ontology: »In order to be in this list [of good ontologies], the ontology must have a documentation page which describes the ontology itself, as well as all the terms defined by the ontology. It must also be used by 2 (verifiable) independent datasets« (W3C 2013). However, it *will* attempt to develop a semi-formal intellectual groundwork for prospective formalization according to specification, which will be scalable for diverse needs. This process does not attempt a complete, thorough portrayal of the Digital Methods domain

beyond the boundaries of the book. Despite the limitation of the collection and interpretation phases to reasonable scientific effort, the ephemerality of web content, web-related studies *and* web-based methods, the constant and rapid transformation of the web itself, prevents from any aspiration of completeness.

1.3 Method

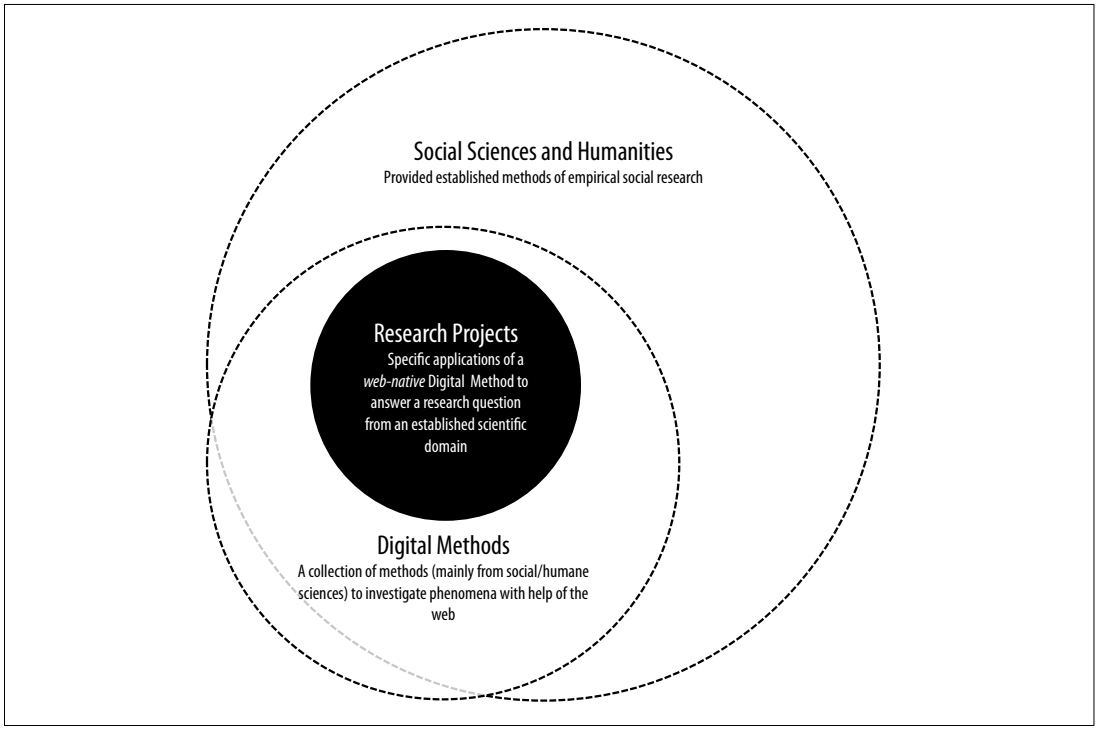

Illustration 1-1: The Fields of Investigation (own illustration)

This paper proposes a formalization supporting comprehensible access to the current state of a research field and a visualization of relationships among the research in this field. More concrete, it attempts to construe and illustrate a knowledge domain that is built up upon three areas, which are interconnected and interdependent. As Illustration 1-1 shows, these three areas are 1) some subset of the huge field of social sciences and humanities, namely those that 2) apply Digital Methods to investigate phenomena with help of the web. Within this intersection between Digital Methods, thus methods that are »born in the web«, and social science and humanities, thus scientific domains that investigate social interaction and human behaviour, lie 3) some research projects that have already been conducted and that used certain Digital Methods. Despite gathering and ordering these projects and their respective methods, the ontology attempts to assign them to the specific scientific domain that is perceived as its scientific »pioneer«

– this can be detached from the Digital Method. The ontology seeks hence for an illustration of all items that sit somewhere in the three sections, and their interconnections.

This paper proposes a stepwise approach to the ontology; starting with defining *what* exactly shall be found (ontology aim), it proceeds with *how* a reliable ontology shall be established (process definition), and how this reliability can be evaluated subsequently (evaluation).

1) Ontology Aim

The right *kind* of formalization has to be found. This paper uses Protégé, a visual editor for ontology based on the Web Ontology Language, widely used for multiple purposes: »Protégé is supported by a strong community of academic, government, and corporate users, who use Protégé to build knowledge-based solutions in areas as diverse as biomedicine, e-commerce, and organizational modelling« (Stanford University 2014a). Besides the desktop application, Stanford University provides a web-based hosting of Protégé to »create, upload, modify and share ontologies for collaborative viewing and editing« (Stanford University 2014b). This might become interesting for future uses of the ontology, when collaboration shall be encouraged. Previous to implementing the ontology within the editor, a frame or guide has to be developed, along which Rogers' book will be scanned for interesting items to integrate in the ontology items will be defined later: What are crucial parameters, both formal and content-wise, for a subsequent usage of the ontology output as a research »assistance tool«? Which taxonomy is able to unite all aspects (layers) of the illustrated knowledge? Which degree of abstraction is appropriate? And how will the various relationships between perceptions in the book and traditional research fields be visualized?

2) Process Definition

The process(es) of data collection, interpretation and processing have to be defined. The utilization of the Web Ontology Language allows for complex concepts to be built up out of simpler concepts (Horridge 2011: 10). This enables the creation of a reversed tree structure with a narrow top consisting of abstract concepts, branching out on sublevels until reaching a widely ramified bottom, in which concrete concepts are displayed. Thus, all studies that belong to the Digital Methods domain can be sorted in a bottom-up structure by starting with the identification of very concrete and unique properties, which are placed in the bottom, and building up higher levels by identifying commonalities and bundling them in superior concepts. Illustration 1-2 shows a simplified structure of the previously introduced examples

of *Google Flu Trends* and *Human Right Types,* as well as possible sibling classes of each level.

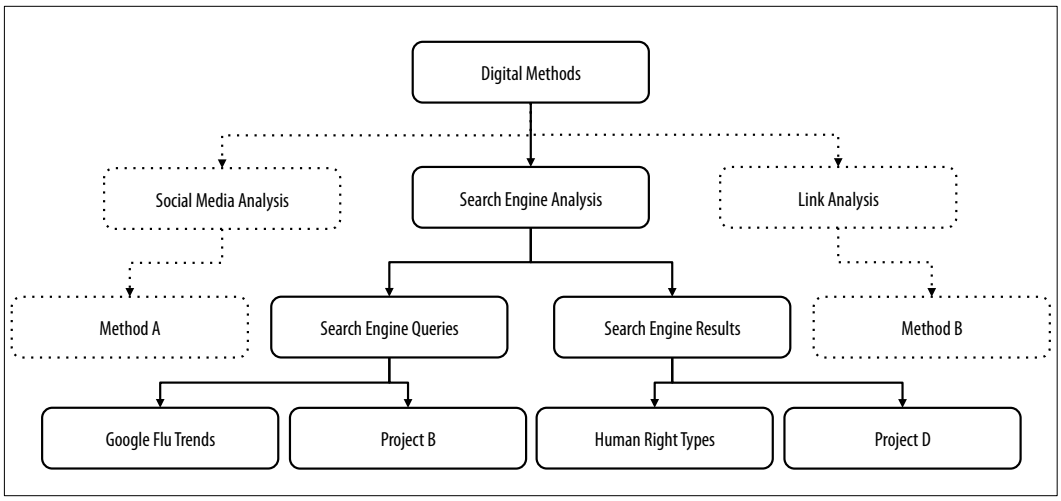

Illustration 1-2: Simplified Reversed Tree Structure of the Digital Methods Ontology (own illustration)

The proposed process is to gather a corpus of studies and evaluate their properties, identify attributes, commonalities and differences among them. Since no superior classification schemes or methodological evaluations are available for this new research field, the sorting logic will be developed inductively along with the identification of entities, and the resulting taxonomy will be evaluated against a second corpus of studies. Due to the empirical approach, the collection phase will need several repetitions. The previously developed scheme will be adjusted along with the processing. What information is relevant? How to prevent relevant information from getting lost in the collection phase?

3) Evaluation

Both the process and the intended result require a thorough evaluation: Were key assumptions of Richard Rogers obtained in the ontology? Did he himself miss crucial parts of the defined research field – and is that showing in the ontology? This question is obviously challenging due to the diversity of scientific interest in the web; nevertheless, a method to evaluate is aimed at. Is the intrinsic logic of the ontology able to thoroughly illustrate a research field, and is it extendable? Concerning the prospective use of the ontology, it is crucial to know whether imaginary usage scenarios of researchers can be completed successfully.

Clearly, from the high-level web science perspective illustrated above, a major part of decisions on the operational level have already been made beforehand to developing the ontology; the definition of Digital Methods, their relevance for web science, the

identification of relevant and irrelevant cases of research and their assertion into clusters of methods is preliminary work that this paper bases upon. The decisions illustrated here are hence residing on a rather operational than strategical level of the »big picture«. The main intellectual work will be composed of the iterative collection and evaluation phases, whereby the composition of the ontology will be based both on gathered and transformed knowledge from the book as well as on empirically derived new knowledge; the latter will mainly be the discovery of coherencies, dependencies and commonalities as a direct improvement, and the generalization of the process as an indirect, perspective feature. Only then will the added value of this research paper for the general (web) scientific community become apparent.

1.4 Structure of this Document

The form of this paper follows the three-dimensional treatment proposed in chapter 1.3, starting with an attempt to answering the questions prompted in (1). Preliminary to working on the ontology, a brief overview shall be given of research that may provide a methodological or epistemological foundation for the Digital Methods ontology. For the eventual utilization of scientific domains within the ontology, a scheme of distinctions and commonalities of related sciences will be drawn, in which all research studies described in the Digital Methods book shall be assorted. The general appearance of the ontology is based on some preliminary considerations on graph appearance and knowledge abstraction. The aggregation process itself (2) will be described, and how the rather random collection of items will be transformed into structural ontology items. Therefore, a discussion about the eligibility of the crucial OWL concepts *classes*, *individuals* and *properties* will be given. The iterative identification of relevant items from the book will be described. The next chapter attempts to answer the questions raised in (3). The ontology shall be analysed for weaknesses in the collection process (*process validity*), the significance and accuracy of the resulting structure (*result reliability*) and the added value of a prospective usage by other researchers (*utilization quality*). For all three challenges, there are a variety of possible instruments for analysis, of respective methods and metrics, available, of which the appropriate ones are identified in the following. Adding to this evaluation, an attempt for interpretation will show whether the statements made in the ontology are generalizable to make statements about the research field as a whole. Concluding, a subsequent discussion will attempt to estimate advantages and disadvantages of the present approach, and identify accomplished tasks as well as questions that might remain unsolved. It will be discussed whether or not the perception of this knowledge

field could be generally enhanced with help of an ontology; if the transformation into a machine-readable meta-language structure contributes to a better understanding of the evolving research field, or to a greater variety of possibilities when utilizing this knowledge for other purposes, and the intrinsic and extrinsic validity of the ontology could be proven, the research question may be perceived as sufficiently solved.

2 Identifying Essential Use

»Don't make me think!« (Krug 2000)

2.1 Essential Use Cases

Preliminary to inducing the ontology, some assumptions about the future usage must be defined to serve as a »guide« through the operative parts. Answering the simple question of »What will the ontology be used for?« will help to specify the actual shape that it shall assume. For this purpose, it is crucial to focus on users and how they interact with the ontology to satisfy certain needs. This is similar to user centered design approaches, with one vital difference. As opposed to software development

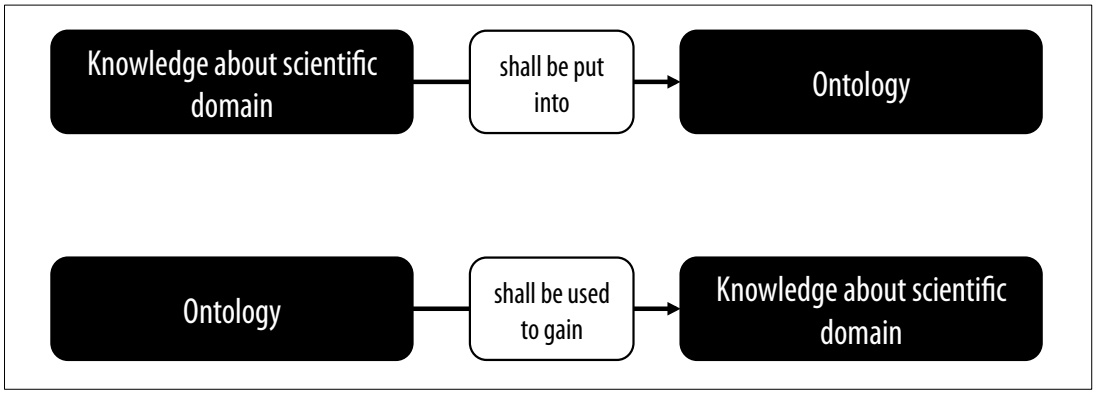

Illustration 2-1: Essential Use Cases (own illustration)

focussing on a *system* that will be *developed presently by person x* and *used afterwards by person y*, where the tasks of the developer differ significantly from any of the tasks of prospective users, working with an ontology requires its conductor to solve the very same tasks that any future user faces: either learn about a knowledge domain by »reading« and exploring the ontology, or expand the knowledge domain by adding new concepts to the ontology. Consequently, only two *essential use cases* are existent for any user: putting something into the ontology, and taking something out of the ontology (see Illustration 2-1). Whereas both activities are intangible[11], the latter is even more

[11] Apart from the intangibility of any digital good, this intangibility refers to the fact that there is no software system of any kind at hand, but solely an ontology that can be experienced in multiple software surroundings, of which Protégé is only one possible application. Neither the interface nor the functionality of ontology editors as such are subject of this research paper, and hence the user interaction refers only to the abstract concept of the ontology itself, represented within any software at user's will.

abstract than the former: What is »put« is literal text, and evokes visible change of the onto-logy, whereas what is »taken« is a construction of previously unknown ideas in the recipient's mind; the outcome is simply *learning* and not visible to other users. Nevertheless, both dimensions of use apply to the conductor as well as to the future user. Consequently, the induction process of the ontology »from the scratch«, which will be described in the following work and afterwards evaluated for its correctness and validity as proposed in chapter 1.3, is in itself already part of checking the use essential use case.

Yet, despite from the essential use, a main difference between conductor and user remains, which concerns the substance. The most salient difference between what the conductor and what the user interacts with is the difference in shape and size between the initial, »empty« ontology and its subsequent complex state. The ontology in its final state has been shaped by its conductor, but as soon as it is released to the public, he loses control over how it is generally used. Which is why the general structure and its self-descriptiveness (and that of the items situated in it) require special attention. To evaluate this, the two essential use cases offer to deduce many *examples* of specific use. Along some *roles* of future users, specific *user stories* can be derived to simulate usage in the most appropriate way.

2.2 Stakeholder Analysis

One is inclined to believe that the »target group« for the ontology is quite small: professional researchers of the web science domain, web practitioners, and more diversified a web-related scientific audience. But, as pointed out in chapter 1.1, despite this rather closed circle, the knowledge about Digital Methods may be beneficial to a much greater variety of scientific professionals. An approach to identifying them is hence necessary.

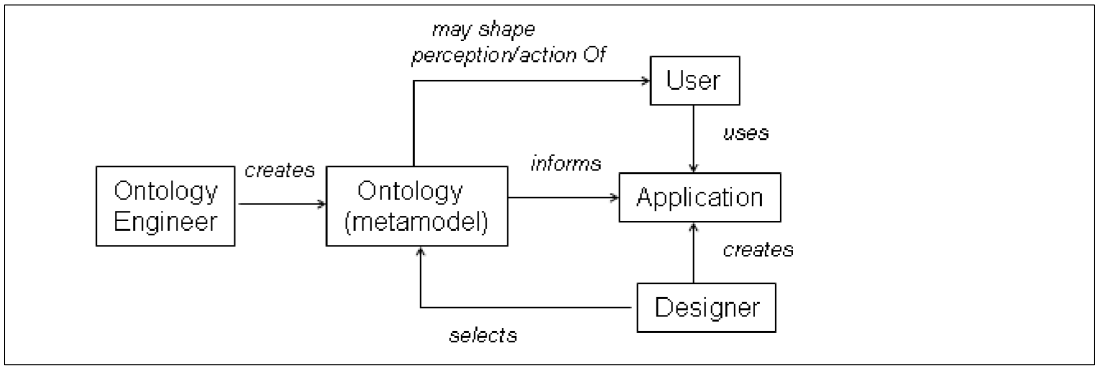

Illustration 2-2: General Framework for Studying Ontological Mediation (Anticoli & Toppano 2013: 25)

In their work about ontology as meta-models in the context of technological mediation, Anticoli & Toppano (2013) identify three functional roles that use the ontology: the ontology engineer, the designer of a web application, and the user. The (simplified) process is that the »Ontological engineer creates an ontology that can be selected by the designer and pushed into the application as a meta-model and emerges when used by the end user« (ibid.). Since their paper focuses on the technological mediation of the ontology, the conceptual background is stressed in which the user is allowed to interpret and use the ontology; the engineer has previously construed the ontology, which denotes the conceptualization of the knowledge domain from the ontological engineer's point of view (interpretation); the designer again uses the ontology as a *meta-model* of the application he wants to develop. This way, perception is shaped by the concepts inherited in the ontology. It was previously stated that the concepts of *engineer* and *user* are not necessarily separate in the context of this paper. Instead, the close resemblance of use also implies a close resemblance of both roles' reception of concepts and relations. To construe exemplary situations of use – or user stories – they can hence be merged into one group: Although they are designed for users, the ontology engineer may find himself in all applications of the user stories. The *designer* that Anticoli & Toppano refer to can be disregarded in the context of the present paper, since his conceptualization of an application draws on the final version of an ontology, and hence on the completion of this work. The user however is conceptualized in four possible shapes in Table 2-1 as *stakeholders.*

Designation	Assumed Interest in Digital Methods	Relation to Ontology
Prospective research conductors with motivation to use web-native data sources	Gain methodological insights and learn about methods and their applications in research projects	Interest in exploring methods and studies resting upon them to derive approaches to own research question
Research conductors with intention to contribute to knowledge domain	Release own work into a professional audience and establish connection to similar projects	Interest in understanding the ontology's underlying structure and find appropriate form of expansion
Research conductors whose project have already been integrated by a third party	Evaluate the correctness of illustrating his approach within ontology	Claim for factual correctness and decent integration of his own thoughts on method and outcome; autonomous correction possibility
Web scientists with intention to evaluate all research in the field of the web	Reuse Digital Methods ontology for a more broadly conceived systematization scheme	Interest in exploring studies and related methods, referential domains of interests

Table 2-1: Stakeholders with interest in the Digital Methods Ontology

All four stakeholders are targeting the Digital Methods ontology as a knowledge representation that can be explored and expanded. From these rather generic groups, concrete user stories can be derived that illustrate the use and hence qualify the requirements that the ontology shall meet in the end. One possible user story has been identified in the example given on page 12: a computer scientist that may build upon the method of a social scientist to answer a different research question. Other user stories can be deduced from the stakeholder analysis.

2.3 User Stories

The user stories can certainly not illustrate all possible usage exhaustively, but exemplary illustrate a variety of intentions for either *essential use case 1* or *essential use case 2*. They can hence be used for the evaluation of the ontology, assuming that if these user stories can be covered with the ontology, the essential use cases can be covered,

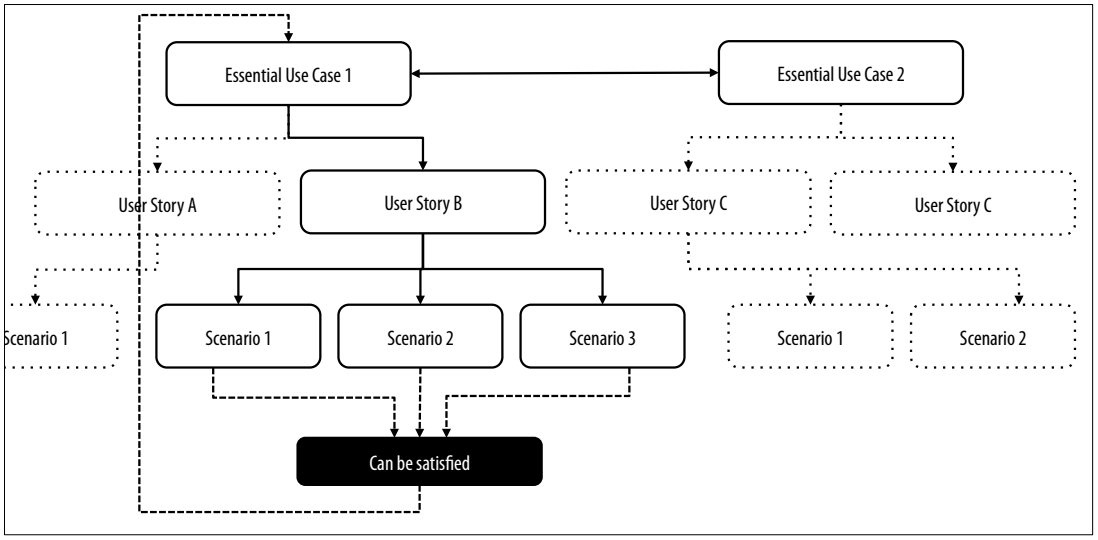

Illustration 2-3: User-focused Evaluation Process from Essential Use Cases To User Stories to Scenarios (own illustration)

which again will demonstrate whether the ontology meets its own essential standards (Illustration 2-3). To evaluate whether or not the goals in the user stories will be met, some *scenarios* of usage will be established in chapter 6.4 for each of the four user stories.

User Story 1: Find research projects that concern search engine usage and its impact on societies, and evaluate the related methods for their ability to be reused for own research project about the political landscape within a language sphere manifesting in search phrases.

Role: Political scientist in research planning phase

Goal: Explore the academic field of Digital Methods and retrieve a set of studies that utilize web-native data of search engine usage for social research. Find out which insights into society are possible with help of search engine user data and how societies are demarcated in this digital data, and get inspiration on how to conduct the research.

Table 2-2: User Story 1 – Retrieve Methods Suitable for Reuse

User Story 2: Explore the ontology to learn about the scientific domains that have been researched with web-native methods so far, and integrate own results of a research project in the domain of cultural studies that was conducted based on ratings on the Internet Movie Database (IMDb).

Role: Cultural scientist in a post-research state

Goal: Explore all items that say something about scientific domains using Digital Methods, understand their relationships and their textures. Identify the appropriate place to integrate the essential information concerning own study: method, size of data set, results, and key data.

Table 2-3: User Story 2 – Integrate Own Findings in Appropriate Location

User Story 3: Scan the ontology for all conductors of studies to find own name, and from own name follow outgoing paths to other information, such as studies by this author, methods used in these studies, questions asked in these studies, motivation to conduct these studies, etc.

Role: Conductor of research project that has been illustrated in Rogers (2013) and in the ontology.

Goal: Retrieve own name and from there discover all information that has been related to it; thoroughly understand all relationships within this information to evaluate whether the information is »right«, and countercheck what information was taken from the book instead of construed in the ontology.

Table 2-4: User Story 3 – Retrieve Information about own Project and Evaluate Correctness

User Story 4: Explore the ontology and comprehend the logic upon which it builds, estimate its significance for the field of web science and reuse it entirely or partly to place it in a broader context.

Role: Web scientist who attempts to classify all web-related research

Goal: Explore parts of the ontology without a defined task and retrieve interesting research projects with attributes, including their intellectual ancestors, their contributors, similar work, etc. Put single findings together as a whole to understand the general field of Digital Methods, discover what it consists of and what not. Demarcate it from other web-related research such as methods to evaluate the web as a medium, and merge both fields into broader sense.

Table 2-5: User Story 4 – Comprehend Knowledge Domain and Reuse it for Broader Context

3 Epistemological and Methodological Foundations

»Because all expressions of human culture are related and interdependent, to gain a real understanding of human society we must have some knowledge of all its major aspects« (Hunt & Colander 2014: 3).

3.1 Introduction

Although, as stated before, there is no framework for the classification of web science research fields, and the scheme will be developed from bottom to top, some initial high-level assumptions will provide means for a reliable empirical process with as few biases as possible. Previous to defining *how* the aggregation and evaluation of ontology items will proceed, it has to be clarified *what* exactly it will consist of – and what not. This is crucial due to the previously mentioned transformation of narration of text in the book into small factual pieces: In the first step of construing the ontology, all narration is broken apart into non-interrelated pieces without a context that defines them. Only in a second step, these pieces are assembled again into something meaningful. The assembly hence needs to be grounded in homogeneous steps for each item. How exactly can 29 pages about *Source Distance* (Rogers 2013: 95-123) be transformed into only three (!) items about this method in the ontology without loosing crucial information? This is only possible if additionally to a thorough process, epistemological and methodological groundings lead through the decomposition and assembly. Whereas the former is important to distinguish important from irrelevant knowledge in the book and will hence support the decomposition, the latter will support a general understanding of the nature of ontologies and hence addresses especially the assembly. Thus and more generally, the following epistemological foundation defines the knowledge area of Digital Methods and demarcates it from similar domains of knowledge, whereupon the methodological foundation will define and demarcate methodological approaches and related scientific domains.

3.2 Epistemological Foundations – Definitions & Differentiation of Domain

When aiming at a thorough illustration of a research field, it appears to be inadequate to focus solely on the Digital Methods descriptions and respective examples of conducted studies provided by Rogers – nor appears the term *Methods* in Digital Methods to be adequate. It is obviously not perceived deficient as such, but possibly misleading in the context of the ontology: The knowledge representation is not focused solely on methods or studies, but rather on the *symbiosis* of research projects, their applied methods and their respective scientific »pioneer«, if existent. »Pioneering« however does not only refer to the methodology on which a research project relies, but may also point to the epistemology in which it is grounded.

3.2.1 Digital Methods are Web-native Methods

As illustrated in chapter 1.1, Digital Methods refer to the amount of web-related research projects that would not exist without the web, be it because they make use of web data in empirical-statistical practices, or use the web as an instrument to use a greater, further afar audience in less time. Whereas both applications point at Digital Methods in the sense of »grounded on digital data«, only the former is web-native in the *proper* sense: The latter is a simple replacement of analogous tools with digital tools. To clarify the circumstances under which a research project falls under the »Digital Methods« definition in the sense of this paper, the term »digital« is an important separator: »Digital Methods provides means distinct from other contemporary approaches to the study of digital material, such as cultural analytics and culturomics, which both make use of the digitized over the natively digital« (Rogers 2013: 204).

Digitized versus digital data is a matter of many definitial attempts and has been subject to change along the evolvement of »The Digital« in culture throughout the years. In the 1960s, a majority of researchers perceived digitized as the transition from analogue to digital data with help of (mechanical or electronic) converters, in the 1970s more and more papers focused on digitalization of communication processes, research in the 1980s experimented with compression of large analogous data sets into digital data, and with the internet as a mainstream phenomenon, parameters of »The Digital«

changed dramatically.[12] Expressions like »Digital Divide« and »Digital Natives« indicate the importance of digitalism in societal and cultural coherencies. Nowadays, »The Digital« appears as a non-defined, almost rather ideological than technical expression, subsuming impact, actions and epistemological states of, within or outside the web. A suitable definition must hence be able to distinct digitized from digital without narrowing the focus to a solely technical level; Rogers himself provides such a distinctive yet holistic definition by a simplification of the parameters:

> »An ontological distinction may be made between the natively digital and the digitized, that is, between the objects, content, devices, and environments that are »born« in the new medium and those that have 'migrated' to it« (Rogers 2013: 19).

3.2.2 The »End of the Virtual« and the Beginning of »Online Groundedness«

The concept of »The Virtual« as a separated space of interaction and existence was the prerequisite of the first phase of internet studies. The perception of separated online and offline culture(s) becomes apparent in examples like the introduction to »Notes Toward a Definition of Cybercommunity«, given by its author Jan Fernback:

> »For those scholars researching the rich terrain of social relations in cyberspace, there are methodological concerns that alert our sensibilities as researchers. How can we apply traditional sociological terms to the patterns of human interaction that develop in the 'bodiless' province of cyberspace?« (Fernback 1999)

Rogers suggests that this distinction is not sufficient due to the tight conjunction of »The Virtual« and »The Real«: »Das 'Reale' wird durch die virtuellen Interaktionen weniger ersetzt als vielmehr ergänzt; diese stimulieren eher reale Interaktionen, als dass sie Isolation und Verzweiflung mit sich bringen würden« (Rogers 2011: 62). This is of utmost importance for the following reasoning of »mapping« the Digital Methods to research practices like those of social sciences and other methodological pioneers, hence to assume that common fundamental principles and preliminary knowledge of researching »The Real«, e.g. of network theories, can be applied to »The Virtual«. Taking this thought further, one could argue not only that methodological foundations are applicable on the web, but also that online data is as valuable as offline data for insights into culture and as a generic source of knowledge about culture. Rogers suggests to call this assumption *Online Groundedness* (Rogers 2013: 19).

---

[12] These findings were made with help of a sample in the IEEE Xplore Digital Library, where the search term »digital« could be applied to a large collection of papers that partly date back to the 1960s. The retrieved papers were sorted chronologically to compare all abstracts.

3.2.3 Medium Specifity

The idea to illustrate the Digital Methods domain with help of the semantic web language OWL evolved due to the nature of the domain of interest (the web) itself; one could argue that it is the most appropriate taxonomical representation of this very domain for several reasons:

1) OWL supports an inductive bottom-up approach, which is essential due to the absence of a generic superior scheme.
2) The representation allows for prospective scales according to the needs of the evolving discipline.
3) OWL as a concept is itself born in the web and therefore follows the medium's very own nature.

This self-referential process of »following the medium« is known to Rogers as *Medium Specifity*:

> »More theoretically, following the medium is a particular form of medium-specific research. Medium specifity is not only how one subdivides disciplinary commitments in media studies according to the primary objects of study: film, radio, television, etc. It also refers to media's ontological distinctiveness, though the means by which the ontologies are built differ« (Rogers 2013: 25).

According to Rogers, it is advisable to follow the medium's (the web's) very own suggestions of dealing with objects like links, threads, algorithmic functionalities or folksonomical sorting:

> »Die Medienspezifizität, die hier gemeint ist, liegt nicht so sehr in McLuhans Beanspruchung der Sinne (…) oder in den Eigenschaften und Befunden anderer Theroetiker. Vielmehr liegt sie in der Methode. Ich habe das an anderer Stelle als 'Web-Epistemologie' beschrieben« (Rogers 2011: 65).

By learning from the medium itself, e.g. about the way search engines prefer certain links over others, researchers may apply the best-suitable method (set). Apparently, Medium Specifity applies to the web as a whole and therefore to the illustration with help of OWL – and serves as yet another confirmation of its usage.

3.3 Methodological Foundations

3.3.1 Methodological Grounding on Empirical Social Research

The Digital Methods ontology is generally challenged by the absence of referential literature. An attempt to providing access to a considerably new and highly transformative domain of knowledge is that of Altmeppen, Weigel & Gebhard, who tried to systematize the domain of communications science with help of an empirical investigation among 835 interviewees (2011: 376). It was established in 2009/2010 to gain insights into the current status of research in communications science. Due to the close embedding of the conductors into the German Communication Association (DGPuK) and the comparably extensive time period of the survey, the research landscape was illustrated pretty much exhaustively. The results show that professionals in this communications field address research in six major areas: public relations, journalism, political communication, media reception and use, media impact and media content (ibid.: 380). Although the project is in fact comparable to this paper in terms of the study object, it differs significantly in terms of groundings, extent and method. The resulting structure of communication science is valuable for considerations about the scientific domain, albeit suffering from the same problem that was illustrated on page 10: It can only cover a snapshot of the field at a specific time, and is not able to acknowledge future changes of the domain. In this case, that effect is even amplified by the research design, which would require another exhaustive survey of domain experts. The approach of developing an ontology for classification appears more suitable in the light of this. Still, the example shows that a taxonomical illustration of a dispersed and transitional knowledge domain is important for the advancement of this field.

A methodological grounding for the context of this work can be derived from empirical social science, which strives for a generalization of observations to make statements about a social context. According to Benninghaus (1998), empirical social research in practice strives for dividing the researched world into *attributes* of *units of analysis*: There may be multiple manifestations of units of analysis, like individuals, cities, nations, as well as multiple different attributes of these units, like interests, income per head, colors of skin; and one attribute may manifest in several units, each with possibly different values. These flexible attributes are consequently called *variables*. Empirical research shall comprise three important tasks:

1) The description of units of analysis and their variables: Summarize observations about (and experience with) a certain object and its variables and represent it.
2) The description of relationships between variables: Strive for implicit correlations between values to become explicit, so that deducing variables of one unit is possible on the base of knowledge about another unit, and insofar predict correlations to reduce the complexity of certain experience values (or research data).
3) A generalization of results: Draw conclusions on the basis of limited knowledge by generalizing certain experiences; previously experienced correlations between variables are often perceived as generalizable for the future, e.g. dark clouds evoking the desire to leave the house with an umbrella. Whereas such generalizations often »fail« in everyday life, empirical research provides measures to estimate the certainty of generalized data with statistical interference (ibid.: 11).

Units, attributes and variables will find their way into this paper by transforming them into OWL items. The challenging generalization of results without statistical interference will be tackled in chapter 7.

3.3.2 Identifying Context: Research Domains with Impact on or Relations to Digital Methods

Additionally to an alternative illustration of Digital Methods and respective research projects, this paper attempts to discover all fields of traditional sciences that are concerned with the interconnected field of web-native research methods. Preliminary to the implementation phase, this requests for a clear vision of the possibly related sciences and their methodological texture. However, a clear distinction between the numerous scientific fields and subordinates that are concerned with human and social interaction (and hence with possible varieties of the Digital Methods) is difficult due to the growth and transformation of disciplines within many years: A room of social interaction is always shaped by its surrounding culture, and so will the research concerned with it have different manifestations. As a consequence, both social science and humanities underlie a continuous transformation of study objects as well as the methodologies.

Additionally to the constant transformation of objects and methodologies that exacerbates the identification of disparities and commonalities, the approving of distinctive definitions of domains depends strongly on the perspective of the approver.

Thus, a distinction made by dictionaries or professionals of *one* field may not necessarily be satisfying for *another* field.

To obtain a clear picture of all relevant domains anyway, a sufficiently credible, universal scheme is desirable upon which as many and diverse experts as possible agree. However, neither surveying experts nor retrieving existing literature can solve this sufficiently due to the one-dimensional perspective of the experts and the resulting lack of consistence in the perception of what would be the definition of »sufficiently credible«. The problem is referred to by Hunt & Colander as »interrelated knowledge«:

> »Because all knowledge is interrelated, there are inevitable problems in defining and cataloging the social sciences. Often, it is difficult to know where one social science ends and another begins. Not only are the individual social sciences interrelated, but the social sciences as a whole body are also related to the natural sciences and the humanities« (Hunt & Colander 2014: 3).

As a result, a workaround is proposed that is based on the idea of Chris Dede's perception of Wikipedia as a collective agreement about knowledge: He proclaims that while traditionally, experts »with substantial credentials in academic fields and disciplines seek new knowledge through formal, evidence-based argumentation, using elaborate methodologies to generate findings and interpretations« (Dede 2008), knowledge in Wikipedia was construed as collective agreement about a description, that it may combine facts with other dimensions of human experience, like values or opinions, and that Wikipedia articles were considered »accurate« when undisputed. Consequently, it appears legitimate to learn from Wikipedia about the distinctive definitions of possibly relevant scientific domains, and use them in the present research paper as long as they appear undisputed. Dispute is conveniently illustrated for readers with a notification in the article's headline, and can be further evaluated on the discussion page. As per elimination process, it can hence be said that every article evoking minor or no discussion, can be relied upon in this paper. Based on the findings, a scheme of scientific areas could be drawn both for the English and German varieties of related research (Illustration 3-1).

In both language spaces, there are a number of scientific domains with a relation to both *Geisteswissenschaften* and *Sozialwissenschaften,* resp. to *Humanities* and *Social Sciences:* cultural studies, communications studies, and anthropology are domains that concern themselves with the individual (humane) as well as the community (social). In Germany, a slightly different group of domains is perceived as in between the dimensions. This may result from the fact that in Germany, there are in fact *three* disciplines: *Geisteswissenschaften, Sozialwissenschaften* and *Geistes-* und *Sozial-*

wissenschaften[13]; a disassociation, e.g. of information science, is even more complicated – and often less desired – than of the English equivalent.

Whereas in some rare cases the ambiguity of branches is due to vague definitions, it is mostly due to the methodological approaches that are dominant in these fields, which often draw upon traditions of both domains; media studies, for example, use a variety of research methods that origin (as they do in general) in communications science; nevertheless, parts of media reception research concerns the impact of mass media on the individual, and does hence belong to humanities according to the scheme.

In both languages, ambiguities between the two disciplines are hence unavoidable. Consequently, a strict distinction between epistemological and methodological relationships of studies or methods to established sciences cannot be drawn. The initial

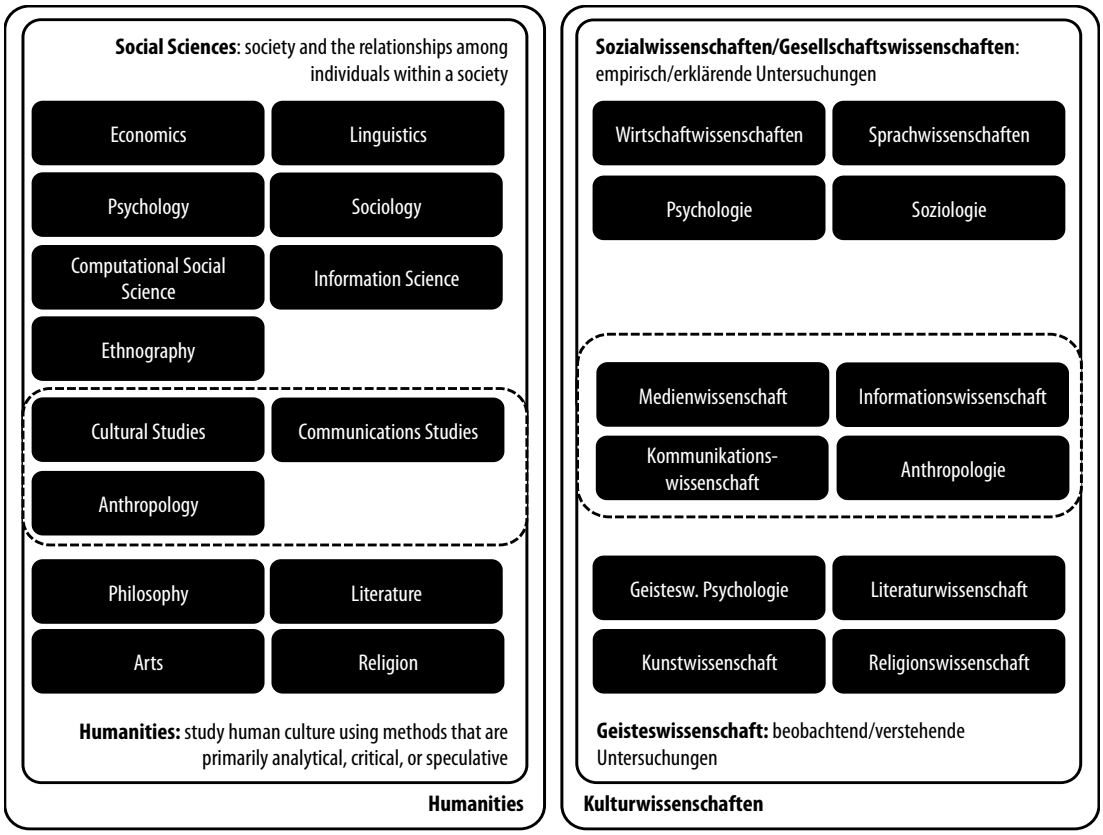

Illustration 3-1: Anthropological Scientific Disciplines Demarcation for English and German Language Spaces, According to Wikipedia (own illustration)[14]

[13] This term is occasionally used to stress the fact that some sciences or concepts are not clearly and distinctly assignable to one or the other domain, but contain issues of both.

[14] Definitions were based on respective Wikipedia articles, cf. Wikipedia (2013b), Wikipedia (2013c), Wikipedia (2014a), Wikipedia (2014b), Wikipedia (2014e), Wikipedia (2014f), Wikipedia (2014g), Wikipedia (2014k).

approach to schematize all scientific domains within the social sciences and the humanities, gather and allocate their respective methods and provide a top-down scheme in which Digital Methods applications »fit«, must be perceived as failed. At this point, it must be assumed that instead of using predefined switches for assigning Digital Methods to established domains, deduced attributes – whether methodological or epistemological – will indicate appropriate positions in particular. Subsequent evaluation is required to understand whether this very approach is generalizable for prospective scale.

3.4 Ontological Suppositions

The ontology that was referred to in several locations of this paper up to this page has not yet been further defined or described. So far, it has only been introduced as some technical concept that will hold knowledge about the Digital Methods in a systematized way. This does in fact conform with the simplest nature of any ontology: »Ontology is a term borrowed from philosophy that refers to the science of describing the kinds of entities in the world and how they are related« (Breitman, Casanova & Truszkowski 2007: 19). What exactly the ontology will be comprised of, and how it will include all entities of the Digital Methods and their relations, is further defined by the language used to create it: OWL. So far, it has been stated that the ontology will be created using the Web Ontology Language (OWL), and that this language is capable of a graph illustration, a taxonomical structure of subordinate and superior classes, machine-readable output, and unlimited abilities of scale. These attributes point at OWL's place of origin (as a W3C Specification) and concomitantly at its primary purpose: »The Semantic Web is a vision for the future of the Web in which information is given explicit meaning, making it easier for machines to automatically process and integrate information available on the Web. The Semantic Web will build on XML's ability to define customized tagging schemes and RDF's flexible approach to representing data« (McGuiness & Harmelen 2004). As of 2012, the description of RDF was supplemented with emphasis on the ability to describe relations: »Ontologies are formalized vocabularies of terms, often covering a specific domain and shared by a community of users. They specify the definitions of terms by describing their relationships with other terms in the ontology« (W3C Working Group 2012).

When it comes to the design of this ontology, one of the main advantages of the foundation of OWL on RDF is that the created knowledge about a domain can be displayed in a graph illustration. The lowest common factor of all RDF statements is that they consist of a subject, predicate and object – an expression that can be displayed

in a »triple«. Instead of using the *OWLViz* plugin, which creates a graph similar to the commonly known RDF triple graphs, it was decided to use , a more sophisticated visualization tool that allows for browsing and individually navigating within the ontology by expanding and collapsing nodes and hovering arrows to see their relationships. This illustration is not fixed, but highly interactive: While browsing through the nodes, every user dynamically creates his individual path through the ontology and thereby establishes an illustration of the Digital Methods domain without having to understand its fundamental construction, since concepts can be experienced without thoroughly understanding their systematized neighbourhood. A user might for example stumble upon a term that he is familiar with, like social network analysis, and explore related concepts, like one certain application of social network analysis in a research project about Mendeley, without having to know the taxonomical position of social network analysis within the Digital Methods categorization. By deducing more knowledge about related projects and methods, he would most probably »unintentionally« reach the state of understanding the systematization as a whole.

In general, the usage of an ontological conceptualization is preferable over other ways of systematization like tabular adjustments, because it puts a primary focus on the relationships between items, which have an important purpose in this context, and it requires no hierarchical prioritization of some individuals over others. Furthermore, a tabular classification scheme would not be possible due to the ambiguity of some projects and the fact that they are not always entirely bound to a specific superior discipline. Rather, what appear most important are their relationships among each other, which is why granularity and a network structure are beneficial. However, what seems like a good solution especially in the construction phase has obvious disadvantages: The original thought of an hierarchical classification scheme, in which levels are comparable one-to-one, will yield for the sake of a network structure with bigger and smaller hubs (representing the superior, abstract levels) and diverse items (representing lower, concrete levels) centring around them. It is hence likely that a rather »arbitrary« seeming structure will evolve. This again is browsable with help of plugins like *OntoGraf*, and hence no insuperable barrier.

Concerning the desired scalability, network structures have no constraints for the prospective inclusion of more items on any abstraction level. However, to establish a thorough understanding of the research field, it is not sufficient to arrange the items in an arbitrary network structure. Rather, it is important to weigh items according to their generality; by means of a weighted structure of abstraction, a taxonomy with several abstraction levels will evolve. This structure along with the infinite growth capacity of any ontology allows for the seamless integration of additional studies *by maintaining* a comprehensive structure. Thus, the process requires to evaluate the necessary abstraction level of items and place them accordingly: It has to be decided for every

item whether it shall reside in a high abstraction level on top of the ontology's reverse tree structure, or on lower abstraction levels in the nestled spaces holding concrete concepts. For instance, if *link analysis* is described as a general approach to investigating social situations with help of hyperlinks, it would necessarily be assigned to higher abstraction levels, because it is predisposed to »carry« concrete *applications* of this technique or branch out in more specific *techniques* of link analysis. Then again, if the text had provided a concrete description of someone having applied a form of link analysis to answer a specific question about a social occurrence, then this would belong into low abstraction levels – to stick with the first example, it would fit *into* the superior, more generic link analysis space.

The discussion about abstraction is necessary in another dimension: Besides the abstraction that concerns the spaces within the ontology, which evolve only after a significant number of items had been integrated and a certain size has been reached, another abstraction concerns the decomposition and new assembly of narration into factual pieces. To give meaning to the so-evolved pieces, both actions have to be applied homogeneously and equally for *all* collected items. This will be eased by the preliminary definition of the ontology's abstraction level: If an ontology engineer knows about the desired degree of abstraction, she knows whether certain OWL functions, like data property assertions, are necessary. Ontologies in formal language can have different degrees of formalization that usually correspond with the context of use; they may also be dependent on the abstraction level of the respective knowledge of a domain in natural language. For instance, a taxonomy of all books available on Amazon would be formal at best, because it could then be reused in the web applications for book retrieval. The data properties could be used to assign one ISBN to every book. As opposed, an ontology of mid-range cars would not need to be formal, since a car model described in would never represent a unique item in the world. Instead, it was a general description of *any* car that has ever been produced in this car series. For the Digital Methods ontology, an abstraction level shall be predefined in order to identify the appropriate degree of formalization in OWL. Parson & Shils (2001: xi) provide a social scientific approach to systematize knowledge with four types of systematization, moving in ascending levels from primitiveness to completeness with respect to the goals of scientific explanation:

1) Ad hoc classificatory systems with »more or less arbitrary classes« (ibid.) of general statements
2) Categorical systems with statements of logical *relationships* among classes
3) Theoretical systems with statements of abstract *laws* or expected outcomes from relationships
4) Empirical-theoretical systems with a specification and explanation of empirical regularities.

Since this paper aims at an empirically derived representation of a research field, in which the previous text format is transformed into an illustration of ideas and their relationships among each other, it aims at developing a categorical system according to this definition. This implies that within the ontology, statements about classes and their relationships will contribute to a categorical system of a research domain. The equivalent of this level definition in OWL is to be found in chapter 4.

4 Implementation

»The scientific community may be thought of as a social system that is organized about a type of cultural interest and commitment, in this case, the maintenance and extension of empirical knowledge«
(Parsons 1967: 157).

4.1 Approach to Induction

The following chapters will describe the iterative process of integrating research projects, related Digital Methods, research domains and any other important concept into a granular taxonomy of unique items and their relationships.

Although the homonymic title of the ontology on the one hand and the referred work by Rogers on the other hand suggest congruent content, only a distinct amount of knowledge from the book is relevant in the context of this work: Despite the disregard of research that is outside the previously given definitions of web-native and digital, the transformation of textual narration of a linear text into factual knowledge pieces will, in a first step, *reduce* the amount of knowledge provided, as illustrated in chapter 3.1. Whereas this is desirable and crucial for a comprehensive illustration, it requires a thoroughly planned transformation of concepts from book items into ontology items, and intermediate reviews of the proceedings so far. Concerning the general knowledge spaces that are transferred from the book into the ontology, the initial scheme that was drawn in the introduction (Illustration 1-1) needs review to tackle an important operational problem: The originally accepted intersection of the three concepts *Digital Methods, research projects in which they were applied,* and *traditional scientific domains in which they can be assigned* prevents, if maintained, the ontology from being meaningful, because the intersected spaces would lead to unambiguity of all inherent concepts – since they would belong to three (overlapping) spaces at the same time. During the subsequent elaboration up to this point, the necessity for certain adjustments manifested.

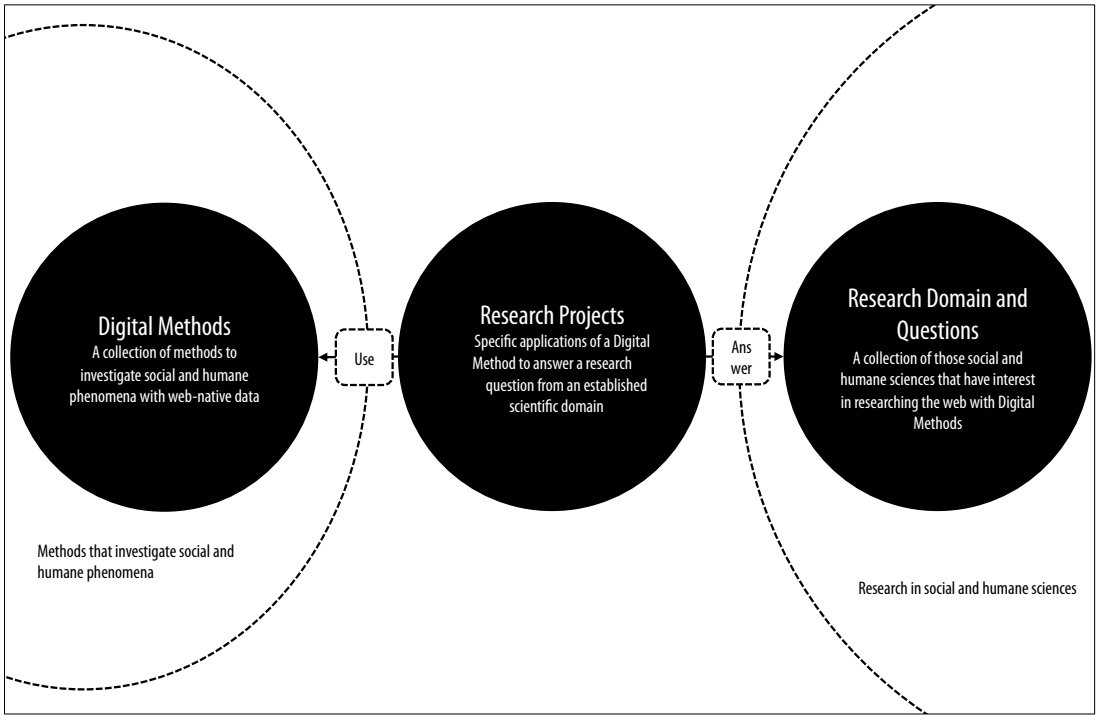

Illustration 4-1: The Triangle of Digital Methods, Research Questions, and Applications in Studies (own illustration)

The provided illustration on page 16 showed three intersected areas, upon which the ontology would be grounded. It is apparent now that the three areas need to be separated from each other in the ontology, and connected only via certain relationships. The new scheme (Illustration 4-1) shows that the originally coherent spaces of *social science & humanities* and *research projects*, as well as the partly coherent space of *Digital Methods*, are dispersed into three absolutely separate spheres. The Digital Methods are now grounded in a space of all »methods that investigate social and humane phenomena«, whereas the research domain and question space is a subset of »research in social and humane sciences«. The semantic difference between the two spheres is marginal, yet their distinction is important for the subsequent collection process. Both spaces are seen as unique and independent and have to be tackled independently as ontology items. Only then will both be represented sufficiently, and their inherent concepts will be unambiguous. The connector of the three spaces will be the projects, which have a relation to both other spheres (they *answer* research questions from a traditional domain, and *use* Digital Methods as a new way of answering). This is not a new thought per se, since it was announced like this in the introduction already, yet the new scheme emphasizes the much stricter distinction.

Since any ontology needs to be construed very closely to the study object – as opposed to applying a generic, top-down framework –, approaching it is a highly inductive process. The ontology structure will hence evolve along the retrieval phase. Attempts of developing a generic framework for the research domains space, in which

items could be assigned in the sense of a decision diagram, must be perceived as invalid – these decisions will hence as well be made ad hoc. Apart from the higher »risk« of inconsistency, the lack of applicable frameworks resulting in this empirical approach provides more flexibility than a methodological set from any other sciences, and by that means also seizes the idea of Rogers to »follow the medium« like addressed before.

Consequently, the initial phase will identify only a rude scheme of collection measures. To accommodate with the flexible and iterative process, this scheme will be adjusted along the integration of additional content. To identify criteria for inclusion, arrangement and connections within the ontology, the initial process scheme attempts to answer five questions, which may be repeated as often as possible:

1) What is valuable content to begin with? Define desired outcome and collect set of items respectively
2) What properties can be derived from the items collected in phase 1? *Extract substance of items*
3) What relationships to additional items do they reveal? Extract additional information and redefine desired outcome
4) Can the properties of 2 be used for the next set of items? *Generalize and adapt for new set*
5) Are the properties collected in 2 and 4 valuable in the sense of providing new insights? Are their relations to other items' properties able to make distinct assertions of anything? *Evaluate findings and iterate*

4.2 Corresponding OWL Concepts

Three major facts are known that will be used in the initial phase: The ontology will provide examples of *studies* in which innovative, web-native methods were utilized; it will contain certain *methods* that are grounded in the online and may or may not have been adapted from foreign domain methodology; it will represent these studies and respective methods in a way that shows their *commonalities* and *differences*. The relationships to traditional scientific domains will be deduced independently in a subsequent step. Since one of the described methods may have infinite applications and is therefore not necessarily unambiguous, the first phase of collecting initial items will concentrate on projects, being the only definitely and undoubtedly unambiguous object: one research project example will become one distinct item, which may be assorted to several methodological concepts. OWL corresponds to this attempt with its three most important concepts (Horridge 2011: 10-12):

1) *Individuals* represent objects in the domain of interest and can therefore be used to represent individual projects, which are unambiguous, distinct from each other, and defined by their relationships to other objects.
2) *Classes* are sets that contain individuals; their description (name) should precisely state the conditions under which an individual can be a member of the group. By creating superclass-subclass hierarchies, the desired taxonomical structure with abstraction levels will evolve. As opposed to individuals, classes are ambiguous by nature; this is why they usually have to be made explicitly disjoint to separate them from one another. In the context of this paper, disjointness will be created only on superlevels, where the knowledge about the class in question is already concrete enough to be sure about this disjointness from other classes. It is for instance legitimate to say that all instances of a *DigitalMethods* class will be disjoint from all instances of a *Conductor* class, but it would at current state not be sufficient to say that all instances in the class *LinkAnalysis* are different from all instances in the class *SearchEngineAnalysis*, because both could hypothetically hold a study about links on a search engine result page.
3) *Properties*[15] are binary relations between individuals; similar to the superclass-subclass-concept, properties can have subproperties and may evolve as a hierarchy. Additionally, two properties may be connected via values like *inverse* (f.i. if property Is*ConductorOf* is the inverse of property *IsConductedBy*, two individuals may be linked together as *A IsConductorOf B* and automatically by the statement's inverse, *B IsConductedBy A*), which means that they can be arranged in a generic scheme, »awaiting« inverse content. This is important for the homogeneity of the collected material (since it may serve as a process control; more on this in in the upcoming chapter) and shall contribute to an efficient workflow.

Given that every single research project may be turned into one individual with several related conditions and that these conditions might as well apply to other studies, the conditions may be turned into classes. Vice versa: Every individual will be assigned to one or several classes that – in total – illustrate the research project's nature. Every individual can be connected to other individuals in one of two ways:

15 Due to the ambivalence of the word properties – on one hand as a concept in OWL, where properties only exist if they create a binary relation between two individuals, and on the other hand as possibly infinite descriptions of conditions of a research project –, the descriptive, second usage of the word is from now on replaced by the word characteristics, features or conditions; whenever properties is used in this paper, it depicts the OWL concept.

a) implicitly through the membership in one class, which means that all individuals in this class share a common value, and
b) by a dedicated definition of relationships to other individuals. These relationships will be identified later.

A commonality of all studies illustrated by Rogers (2103) is that they already have names assigned or that names can easily be derived from the description. The research examples given are hence well suitable to be transformed into the individuals of the ontology: They are each specific instances of the classes that hold them and are clearly distinguishable from one another, although they have enough similarities to be grouped together into the same classes.

When using the ontology, two general and competing interests can be identified from the essential use cases (chapter 2.1): One might want to discover and therewith understand the Digital Methods research field as a whole, or one might want to study specific details of one research project and discover all related concepts of this very project. The first intention requires the ontology to contain abstract, high-level classes – as many as necessary, but as few as possible to prevent from unnecessary distraction. As opposed, the latter requires the ontology to contain detailed individual information, especially concerning the properties showing relationships among individuals. The solution to this perceived paradox is to strive for very few items on higher levels, but much more granularity on lower levels. These lower-level properties and individuals will contain more sophisticated information and will most probably be of a higher amount. The general ontology shape is hence rather thin in superior regions and »broadens up« in lower levels, creating a reversed tree structure, as illustrated already in Illustration 1-2. The superior classes will be defined beforehand to the creation to »frame« the subsequent collection, whereas the low-level information will be deduced ad hoc.

4.2.1 Top-level Classes

Whenever a research project is conducted, one would not apply any method as a self-purpose. Instead, one would like to find an answer to a question posed beforehand. Hence, additionally to the three introduced factors *projects, methods* and the *relationships* between them, the next important finding about any research project is the *domain of interest* in which it belongs. Following Illustration 4-1, this interest is always grounded in some traditional domain of social sciences and humanities. Given that this paper perceives web science as an independent domain and the Digital Methods as a contribution to prove this, it might appear superfluous to establish traditional sciences as a whole independent space. But up to this point, the Digital

Methods are not yet established as a research area that exists of its own accord; research methods and experiences are mainly adapted from traditional research, and again transform these research methods now. These interrelations have to be made accessible to research in general.

To find the appropriate item for this, it is once again advisable to strive for *Medium Specifity* – or to »follow the medium« (the medium being the domain of interest): Since ontologies attempt to describe *knowledge*, *questions* are perceived best suitable to represent domains on lower abstraction levels for two reasons.

Firstly, questions are a convenient way to keep the ontology scalable due to the fact that they may arise at any time: It is always possible to ask another question based on the previous one or based on superior clusters that hold them. Secondly and more importantly, processing questions that were answered by researchers is less »risky« than sorting according to precise facts about the foreign domain or any other differentiator, because they always exist no matter what the nature of the research project was. Additionally, questions can be construed by the ontology designer independently from the degree of sophistication of research project descriptions provided. If other characteristics would serve as differentiators and these were not be given by the authors of certain studies, and could also not be deduced by the ontology engineer, these projects would lack parts of the description, and the comparability to other projects as well as the overall meaningfulness of the ontology would suffer.

Hence, by assigning research projects to questions that they are supposed to answer, the ontology will be more precise than by grouping after any other characteristic, and comparability easier to establish.

Concluding this and the previous section, the top-level classes identified so far are *DigitalMethods, ResearchDomain* and *ResearchProject,* whereby research domains will appear in the form of research questions that one specific project answers by applying a certain Digital Method, but which are not necessarily asked *only* in the coherencies of Digital Methods; some may have been there much longer than the web, and applying Digital Methods is yet another attempt of answering them. For example, a research project about the hyperlink structure of political or near-political organizations may be evaluated to find patterns of associations. This would answer a research question from the domain of political science with help of a Digital Method, but the question itself, the desire to gain insights into the motivation of political and near-political organizations to refer to each other or not, is perhaps much older.

4.2.2 Top-level Properties

During the first iteration, after collecting the first sample, it becomes apparent that the previously assumed integration of methods and other characteristics as several classes of one research project (individual) is inappropriate. Not only is it much less significant to connect two individuals of the same class (or in classes with the same superclass) via properties, it is also illegitimate to assign certain individuals to *all* classes in question (of different superclasses). Furthermore, it is not possible to state that an individual has a specific relationship to the class it is in; the only statement that connects both entities is *HasIndividual* resp. *IsIndividualOf.* A simple example will illustrate the logical problem: The research project *GoogleFluTrends*, the first example of projects given (Rogers 2013: 4), was originally defined to belong to the classes *ResearchProject, QueryLogAnalysis* as a subclass of *DigitalMethods* (the superclass subsuming and sorting all described methods), and *CulturalAnthropology* as a subclass of *SocialResearch* in the *ResearchDomain*-superclass. Literally, this would mean that the project was an *area* of social research, and that it was a *method* of investigating the web; none of this is obviously true for a single research project. Rather, it needs to be made clear which *connection* Google Flu Trends has to methods on the one hand and research domains on the other: It *uses* some kind of method to *answer* some research question in the field of some research domain. Object properties become crucial.

Consequently, preliminary assumptions can be deduced about the high-level relationships between individuals of the three superclasses *DigitalMethod*, *ResearchDomain* and *ResearchProject*: *Answers* is the general connection between a specific research project and a research domain, because every research project attempts to answer a research question arising within a certain research field. The rather generic object property *Answers* will be a superior property subsuming all relations from an individual of the class *ResearchProject* to an individual of the class *ResearchDomain* (resp. its subdomain). Hence, subordinate properties will explain the *specific* solution proposed in one project to answer its research question. It is not a description of the method, but rather the unique characteristic of every research project, by which means the research question is answered.

In OWL, constraints can be defined for specific properties, called *domain/range restrictions*. They define relations of all individuals of one class a (domain) to all individuals of another class b (range), stating that any individual from one domain has the property x to any of the individuals in the range. In the present case, *ResearchProject* will be defined as the domain of the property *Answers,* whereas *ResearchDomain* is the range; this means that *Answers* should always connect one individual of the class *ResearchProject* to at least one individual of the class *ResearchDomain.* This would automatically create a statement for the inverse of

Answers, if *IsAnsweredIn* was established*:* All individuals of *ResearchDomain* would automatically be labelled as *IsAnsweredIn* an individual of the class *ResearchProject.* Certainly, the domain/range restriction does not replace manual work; it is solely a control mechanism for the researcher – a so-called *reasoner*[16] detects the illustrated deviations from the scheme and displays them. That way, it also detects »false« occurrences. More on the usage of domain/range restrictions for error detection will be provided in chapter 6.1. The domain/range idea was initially used for several restrictions of the present ontology, but most of them were dismissed in the following because they complicated the process more than supporting it.

Another toplevel property concerns the relation of research projects to their respective Digital Method: *utilize* will hold every specification of the relationship of an individual of the class *ResearchProject* to an individual of the class *DigitalMethod*; these properties will specify the method that is applied for a specific study. Like *Answers,* the rather generic property is further divided into subproperties, so that every research project is connected to a method via a dedicated subproperty of *Utilize.*

As shown in Illustration 4-2, the individual of the *DigitalMethod* class that *IsUtilizedFor* a specific research project is denominated by an inverse; for the process this means it is not required to specifically name this backwards-relation – it will be created automatically by the reasoner. This holds true for all classes in the proximity of

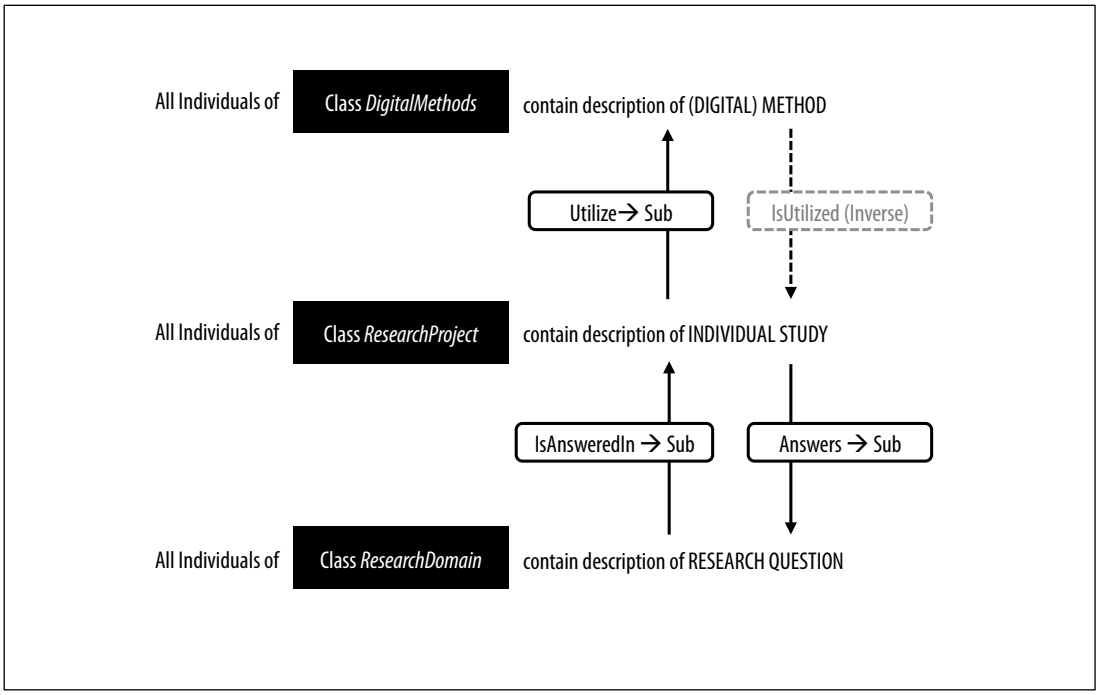

Illustration 4-2: Necessary Relationships (Properties) of Superclasses (own illustration)

[16] This paper uses FaCT++, a reasoner based on C++ for advanced portability, developed at University of Manchester (for further information, visit http://owl.man.ac.uk/factplusplus/).

a *ResearchProject* individual that were deduced subsequently. Some of these relations are already incorporated in Illustration 4-2, others followed subsequently. The only exception from this concept of inverse properties is *Answers* resp. the *not* inverse property *IsAnsweredIn*, which is due to the fact that *IsAnsweredIn* contains several subproperties: It became apparent during the first collection phase that an important finding about any project is on which data it is grounded: Was the user data collected concurrent to usage, or subsequently? Is it grounded in a combination of offline and online, or solely on web-native data? Is the project using repurposed data that has already been existing, or a separate collection of dedicated data for the research project? Additionally, these classes clarify whether the collection was done manually, by automated web scraping technologies (dormant data) or by simulation of usage (ephemeral data). This will provide an illustration of the foundation of data for studies in the Digital Methods research field. The requirement for this distinctive hierarchy becomes apparent in the subsequent exemplary collection.

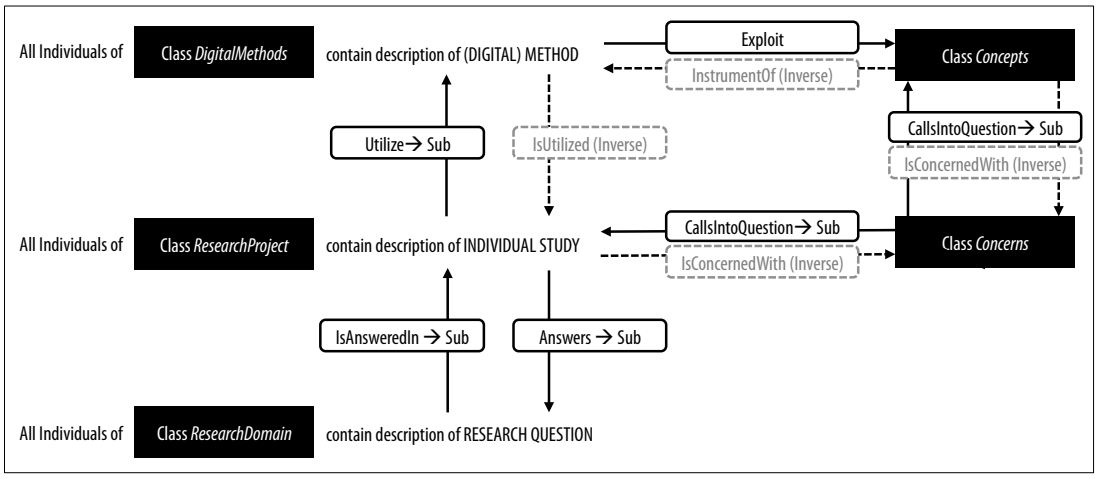

Illustration 4-3: Necessary Relationships (Properties) of Superclasses Extended (own illustration)

Subsequently, two more superclasses were added, namely *Concepts* and *Concerns*, the latter being important for a prospective understanding of the scientific domain that evolves. Criticism can either be applied in a methodological sense (method deployed in the research project) or institutional (research project *reveals* reasons for criticism, e.g. censorship). Chapter 4.4 will explain why they were necessary.

4.3 Exemplary Initial Collection

To illustrate all decisions made and challenges approached so far, the following example will illustrate the first collection phase along research projects. The first research project is mentioned in the introduction (Rogers 2013: 4) to illustrate the

distinctive character of a Digital Method: *Google Flu Trends* stands as an example for »a classic and teachable case of thinking through the availability of natively digital objects (...) and repurposing engine results for social research« (Horridge 2011: 10-12). As explained previously, a specific research project is unique and hence transformed into an individual. The unique name assigned to the individual is *GoogleFluTrends*. Certain possibly relevant statements can be deduced about this research project:

a) It was established in 2007/2008.
b) The number of participants (50 Million search queries from 2003-2008) exceeded by far the number of participants in classical, empirical social research.
c) Results are grounded in the comparison with other, not natively digital data (of US Centers for Disease Control).
d) The method was use existing data – the search engine queries and locations of these queries – and repurpose it for social research to gain insight into flu occurrences.
e) The (deduced) research question is: What do search engine events tell about the real life of people regarding a specific domain, e.g. diseases?
f) The research project was conducted by Google.org, the self-proclaimed non-commercial initiative of Google Inc.
g) The scientific domain related is social science, more specifically the branch of cultural anthropology. This is classified by Rogers and does not require a new establishment of a branch.
h) The related, equivalent (offline) method of cultural anthropology is field studies.

More statements would be possible especially regarding implementation, interpretation and results. Since the ontology does not attempt to be a *complete* guide through all attributes of studies, but rather to show interconnections between several studies, it is important to identify commonalities to other items and to distinguish crucial from additional and irrelevant information. Two simple but crucial statements are those of conductor and year of origin. This information will be included in the respective superclass. When comparing with the user stories defined in chapter 2.3, it becomes apparent that a simple collection of conductors is not sophisticated enough; a more detailed distinction is required. Besides the name, one would probably like to know how big the research team was, what branch the research team is assigned to and what interest (commercial or non-commercial) was behind the project. The latter is important for every research project to provide means for evaluating the credibility of results, and the team size might be interesting for someone who attempts to conduct similar studies. As a result, the *Conductor* class is divided into commercial vs. educational-scientific background (journalistic and artist backgrounds were added

later) and into institution vs. single person classes. This covers the essential information regarding the research's external circumstances. Based on similar considerations, the class *YearOfOrigin* was moved into the superclass *TimeFrames,* which also holds the class *NonNumericTimes.* These were important for some conceptual considerations about the web, e.g. to describe the time in which the web has been seen as a »virtual space« separated from the offline world.

The research question is transformed into an individual of the class *ResearchDomain.* After several iterations, it was found that the class holding a specific question as an individual should sometimes itself be a question; this holds true for the current example as well. The research question posed in this study, »What do search engine queries tell about society?« is partly answering the broader question, »What do natively digital objects tell about society?«, which again is part of the domain of cultural anthropology as a part of social research. The superior question is necessary to provide a location for other projects with a similar intention.

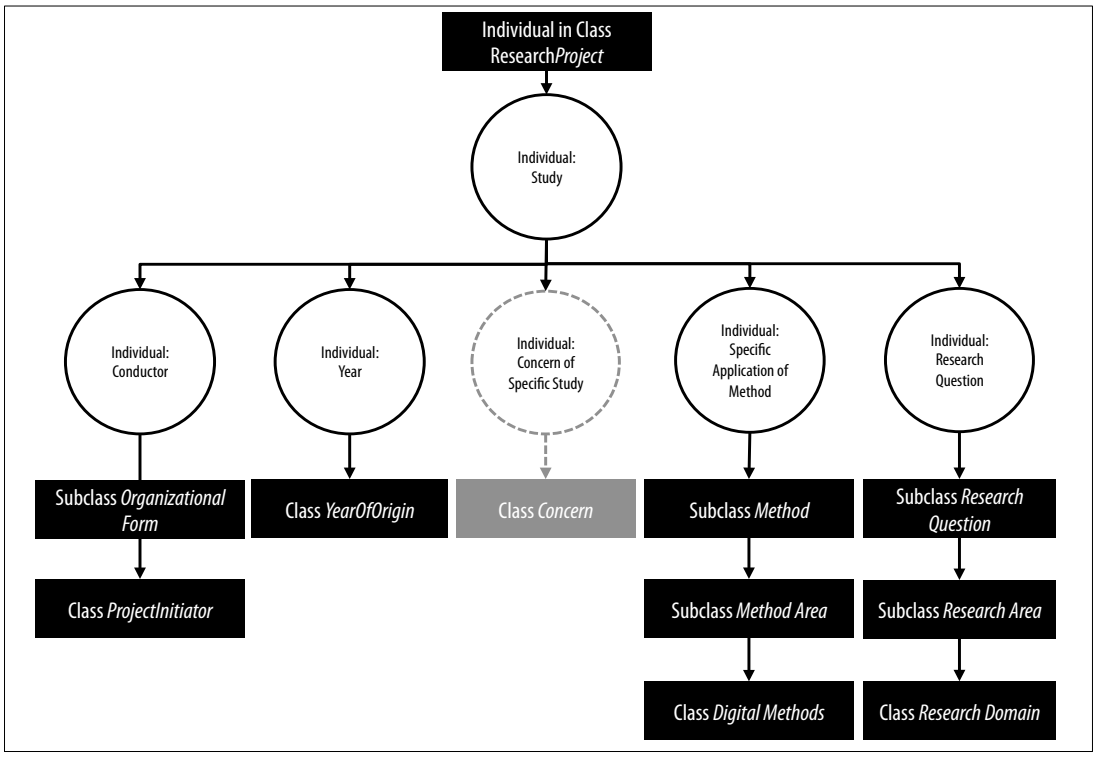

Illustration 4-4: Ontology Tree Structure of Collection Process (own illustration)

The collected information about the method was transformed into its equivalent in the diction of Digital Methods, since only its impact on and relation to the Digital Methods research field is relevant. Hence, the respective Digital Method, *Query Log Analysis*, is reduced to a name of an individual, whereas the method was dismissed in the first step. It was found that refraining from this information led to a significant decrease of knowledge, which is why what was previously collected as a method, was subsequently

included as a property: All *IsAnsweredIn* subproperties are obliged to distinguish offline from online data bases and subsequent from concurrent collection. This subproperty hence illustrates the handling of research data in general.

After adding the necessary additions that were revealed along the first object, the superclass taxonomy up to this point was illustrated in another scheme (Illustration 4-4). It shows that there are five areas in the top level, of which *DigitalMethods* and *ResearchDomain* contain the highest sophistication, manifesting in the greatest subclass dispersion. The integration of new items can now follow a structured process: The research projects as the most concrete possible unit will be the centre of every new set, which is why it will always serve as a starting point, from which all other properties and individuals are assigned to the remaining classes.

4.4 Additional Problems Solved

One major challenge of using an ontology language is that knowledge will be systematized as small, factual instances, as pointed out on page 37. Whereas this is a huge advantage in terms of comprehensibility, some information might get lost, since certain deductions are reserved for human logic: While reading a book, a human is able to identify certain coherencies and make certain assumptions on the foundation of his personal experience and cultural semiotics. He might for instance decide about the integrity of an institution based on the previous experience he has had with it, and he might know about similarities among projects because of their location within the same section of a linear book. In the formalized Ontology, this knowledge has to be made explicit. It was therefore decided to establish another superclass called *Concerns*, in which methodological or entrepreneurial criticism is displayed and assigned to studies. The previously explained distinctive subclasses of *Conductor* (commercial versus educational/scientific) also arose from these considerations.

Another problem is the inability of the formalization to »know« about the proximity of ideas of two objects if it was not explicitly stated, which is illustrated with help of another example: It has been found that there is a research project of the evolving of a page illustrated with help of a *Screencast Documentary*; a movie of changes, adjustments and possible discussion of a specific page. The related project (as an individual) is *GoogleAndThePoliticsOfTabs*, originally introduced on page 16 (Rogers 2013: 16). Now, the next time it is mentioned is on page 68 (ibid.), immediately after introducing a very similar project: *HeavyMetalUmlaut*, the story of the evolvement and professionalization of a Wikipedia article of special interest (Rogers 2013: 68). From human experience with text (subheadings, same location, flow of argumentation), it is

obvious that both studies are interconnected by certain very similar, but not identical parameters or lobsters. It is important to find whether these similarities are made explicit somewhere within the ontology. *GoogleAndThePoliticsOfTabs* attempts to answer the question *HowDoesGoogleWeigh-AlgorithmicOverHumanCataloguing*, which is an individual part of the class *HowDoesMediaPerceptionChangeOverTime*. The *HeavyMetalUmlaut* project on the other hand is classified as answering the question: *HowAreWikipediaArticlesProfessionalizedOverTime*, being an individual of the class *HowDoesTheWebChangeOverTime*. Both classes are part of the superclass *HistoriographicalWebAnalysis* because they make use of the web or a website as an archived object. The proximity of both studies, deducable in fact without further thinking in human logic, is hence established via their mutual superior class, which goes up to the superclass *MediaStudies* of the knowledge area of *CommunicationsScience* as part of *ResearchDomain*. Vice versa: If someone would look up the content of media studies in relation to Digital Methods, he would find that both studies attempt to answer a question in the area of the web as an object of historiographical investigation. Similarly, from the Digital Methods point of view, *GoogleAndThePoliticsOfTabs Utilizes* a method called *DocumentationOfSingleSiteHistory* from the area of *ScreencastDocumentaries*, specifically *HistoricalSiteAnalysis*. Now, *HeavyMetalUmlaut*, utilizing *IllustrateEvolutionAndProfessionalizationOfWikipediaEntry*, which is part of the class *TimelapsePhotography*, is in the very same superior class: *ScreencastDocumentaries*. The proximity is established twice in this case, although it may have been possible to investigate two similar research questions with distinctive Digital Methods. Concluding, one might find that the lower the class level is in which two individuals meet, the more apparent is their similarity. If they meet in two different classes in more than one superclass, as in *DigitalMethods* and *ResearchDomain,* this might as well accelerate the feasibility of similarities in two studies.

In another case, no connection is visible trough either method or domain and a workaround has to be established: The investigation of Google search results for the term »terrorism« points at the same offline occurrence (9/11) as the Whitehouse.gov Issue list, but because the first is assigned to *ComparativeMediaAnalysis* and the latter is *DocumentationOfSingleSiteHistory*, they differ not only in the method, but also in the domain. Both individuals were hence connected to each other via the property *HasSimilarTopic*. Another remedy is the property *IsAdvancementOf*, which connects two studies of the same conductor using the same method and answering the same research question – with the difference that they have been conducted successively and one builds upon the other. In a future use of the ontology, these properties can link together all kinds of individuals, and their proximity could be made machine-readable via a universal specification.

5 Results

» 'I thought you didn't like Facebook anymore?' 'Don't be silly. I'm a fan of anything that tries to replace actual human contact' « (Sheldon Cooper in The Big Bang Theory, Season 5, Episode 10).

5.1 The General Structure of the Digital Methods Domain

All in all, 69 research projects and 39 forms of Digital Methods have initially been identified as relevant from the book (Rogers 2013) during the inspection, and marked as such.

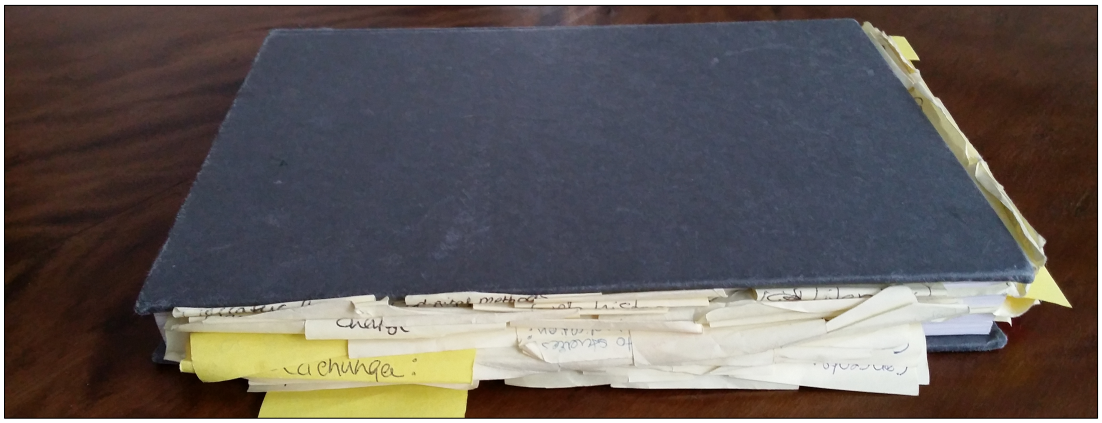

Illustration 5-1: The Digital Methods Book After Inspection (photography)

These were in a second step filtered after methods that do not fall under the previous definitions or were perceived irrelevant for other reasons, methods that were abstractly described, but not provided with specific applications in research projects, and research projects that had deficient descriptions and could not be found in other sources; all described items were dismissed. Certainly, the major part of dismissals was due to duplicates. After cleansing, 31 individuals of the *ResearchProject* class remained. Beginning with these 31 items, the other classes were built up based on the process described in the previous chapter.

The whole Digital Methods ontology is shown as a network structure in Illustration 5-2; it becomes already apparent that the clarity of the structure does in fact improve quick comprehension of the knowledge domain in total. It is now possible to navigate through all items, discover their neighbourhood and show interrelations of concepts by clicking.

As supposed, the Digital Methods ontology finally resulted in seven superclasses, from which several subclasses and individuals depart. As expectable, the classes *Concerns* and *Concepts* are rather small because they were optional and hold only additional information. The classes *DigitalMethods*, *ResearchDomain* and *Research-Project* as well as *ProjectInitiator* and *TimeFrames* build up the core knowledge about the Digital Methods domain. The *TimeFrames* and *ProjectInitiator* classes are described further in chapter 7.2.5 and chapter 7.2.6 – the remaining three mandatory classes will be described briefly in the following.

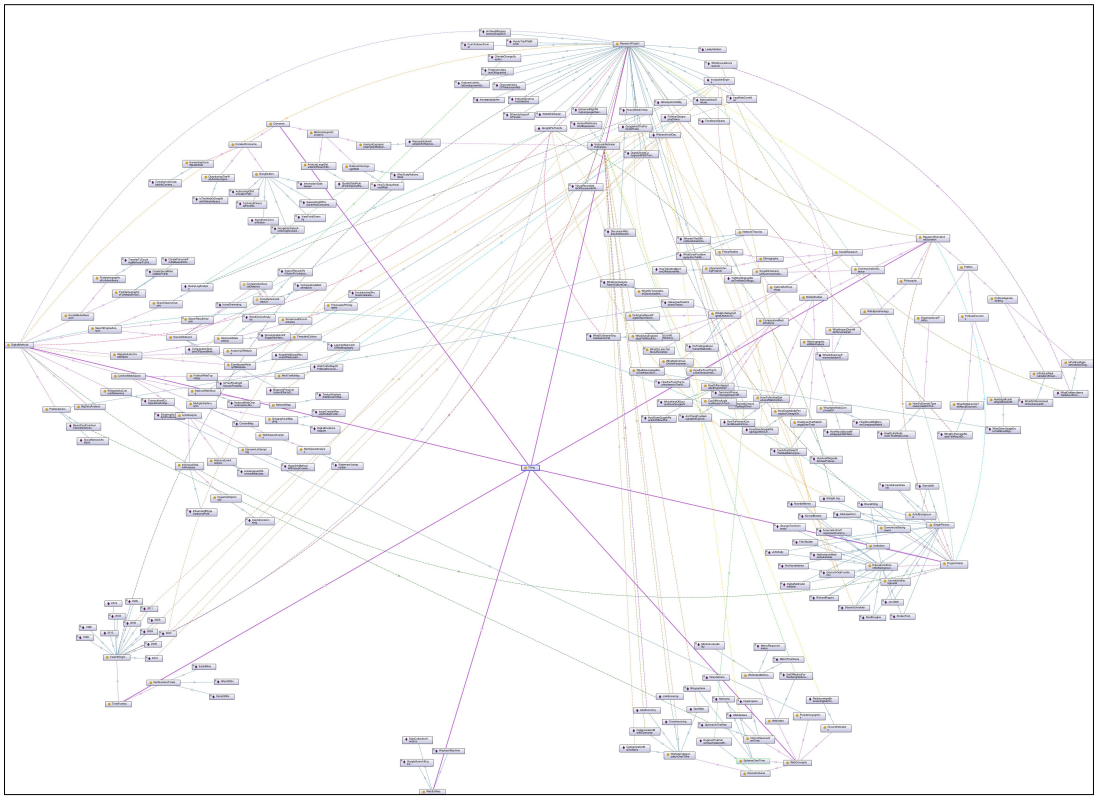

Illustration 5-2: Map of all Ontology Items (Protégé export)

Using Protégé, all content is available in the form of »interactive« lists. Interactive means that by clicking, any relation of one item (class or individual or property) to other items, resp. their usage within the ontology, can be shown. As soon as one launches the reasoner, additional implicit information is made explicit through its visual appearance, as the yellow lines in Illustration 6-2 show. By that means, the entire research field can be experienced in an explorative journey. Alternatively, a visualization tool can be used. With help of *Ontograf*, all items and their relationship are illustrated in a network structure with nodes (for individuals and classes), arrows (for properties among them) and »Tooltips« (for a summary of all characteristics of one node). Again, the approach to reception is explorative: The network structure builds up »from the scratch« during usage and expands and collapses nodes in real-time, as

opposed to providing a frozen illustration of the entire field. Here lies one of the main advantages of using ontologies for complex knowledge constructs: Apart from reducing complexity on the visible canvas, which contributes to a better perception, the visualization tool is able to react on the very nature of different classes, and the single threads are structured just as needed: Whereas the ResearchProject class – consisting solely of unordered, unprioritized individuals – can obtain the form of a simple »swarm« (Illustration 5-3), the complexity of the class *DigitalMethods* – consisting of several subclasses with deeper nestling each, but of which none is prioritized over the other – can be acknowledged by a network structure with one centre (Illustration 5-4), and the *ResearchDomain* class – in which taxonomical structures are important for understanding – can appear highly structured into a lateral tree structure (Illustration 5-5).

The three sub-structures can therefore be described independently, or examined as a whole. In the following, they will be described briefly.

5.1.1 Results in the Research Project Thread

The superclass *ResearchProject* holds a total of 31 individuals, which means that 31 single studies have been discovered in the book that fall under the definition of a research study using web-native data, as was described above. In Illustration 5-3, it is shown that as opposed to every other superclass, the *ResearchProject* class is not sub-

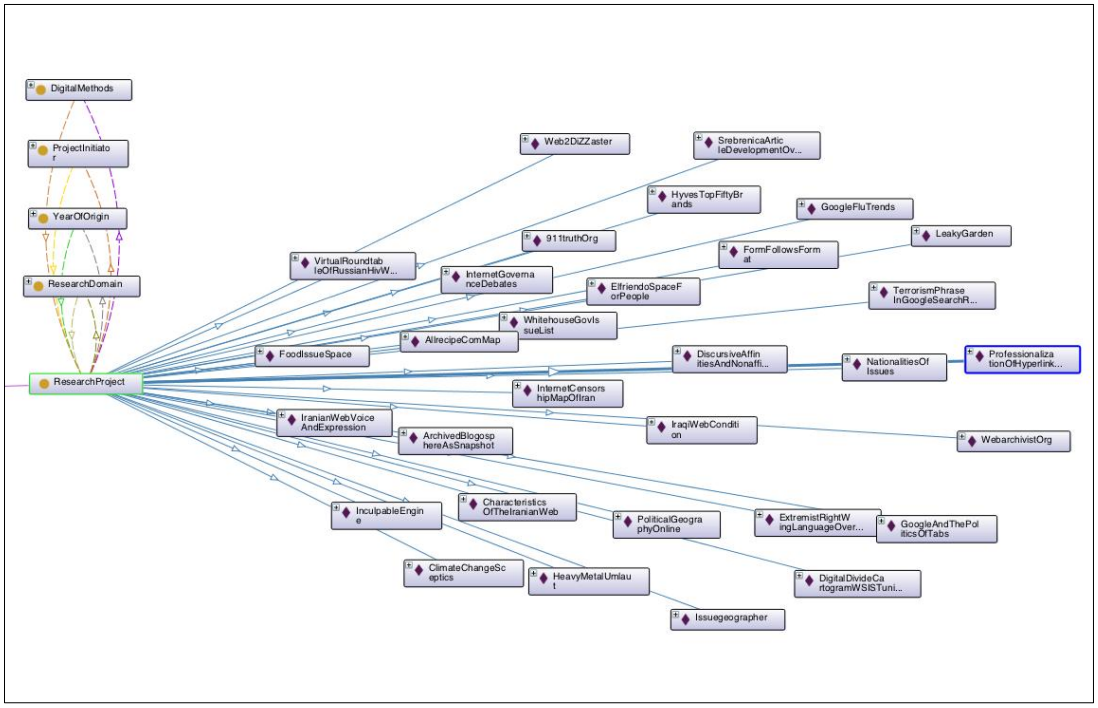

Illustration 5-3: Neighbourhood of the *ResearchProject* class and all contained individuals (Protégé export)

divided into further nuances. The reason is that all contained research projects are defined by their surroundings, thus by their neighbourhood that is construed by the sum of the respective Digital Method, the respective scientific domain, year of origin, conductor and, optionally, related concerns and concepts.

5.1.2 Results in the Digital Methods Thread

The thread *DigitalMethods* also contains 31 individuals – although as a result of coincidence. Since some *DigitalMethods* individual may define one *or more* projects (if two projects use the same Digital Method), the total numbers within the threads do not necessarily correlate. It has to be said, though, that although a difference in these numbers would not be classified as an error, divergence must in fact stay a rare case due to the very concrete fit of *DigitalMethods* individuals to the *ResearchProject* individuals. The class itself is dispersed into six subclasses, in which methodological areas are further described, as shown in Illustration 5-4. According to this, the Digital Methods currently consist of methods in the fields of

- »Link and web space«, hence investigations with help of links, link maps and network maps,
- »Search engine analysis« with search results or search phrases as data, hence search-engine-wise or user-wise methods, as well as source distance, a special form of analysing the attention that certain stories receive in search engines,
- »Social media research«, or »Postdemographics«, a term coined by Rogers in the context of these methods,
- »Website as archived object«, which contains analyses of single websites as closed objects as well as historical site analyses, hence website developments overtime,
- »Wikipedia as cultural reference", a method to study Wikipedia for insights into culture.

The sixth class is not a description of methods, but instead holds »Predecessors« of methods. On the »ground« levels, every class contains at least one individual that is related to another individual of the class *ResearchProject* via one dedicated sub-property of *Utilizes / IsUtilizedFor.*

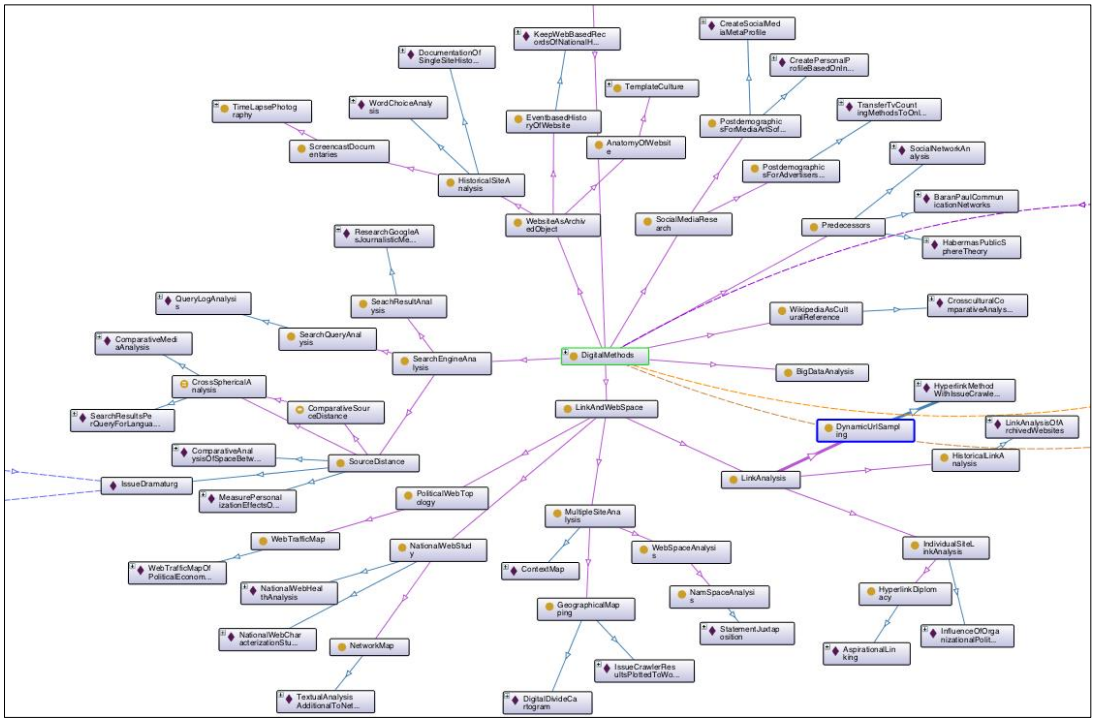

Illustration 5-4: The *DigitalMethods* Class Thread with all Contained Entities (Protégé export)

5.1.3 Results in the Research Domain Thread

The thread *ResearchDomain* holds 31 individuals: As opposed to the *DigitalMethods* class, this class *is* correlating with the *ResearchProject* class, since every research project asks (resp. answers) exactly one question. Three classes are directly subordinate to the *ResearchDomain* class (Illustration 5-5): *SocialResearch* and *Politics* as well as *Philosophy*, with *SocialResearch* holding by far the most entities – as was expectable, given that the web is a social interactive space. Moreover, it must be surprising to have classes outside social research at all. In fact, this is more related to the discordance (even within Wikipedia) about the distinctive attributes of the three domains as illustrated in chapter 6.1.

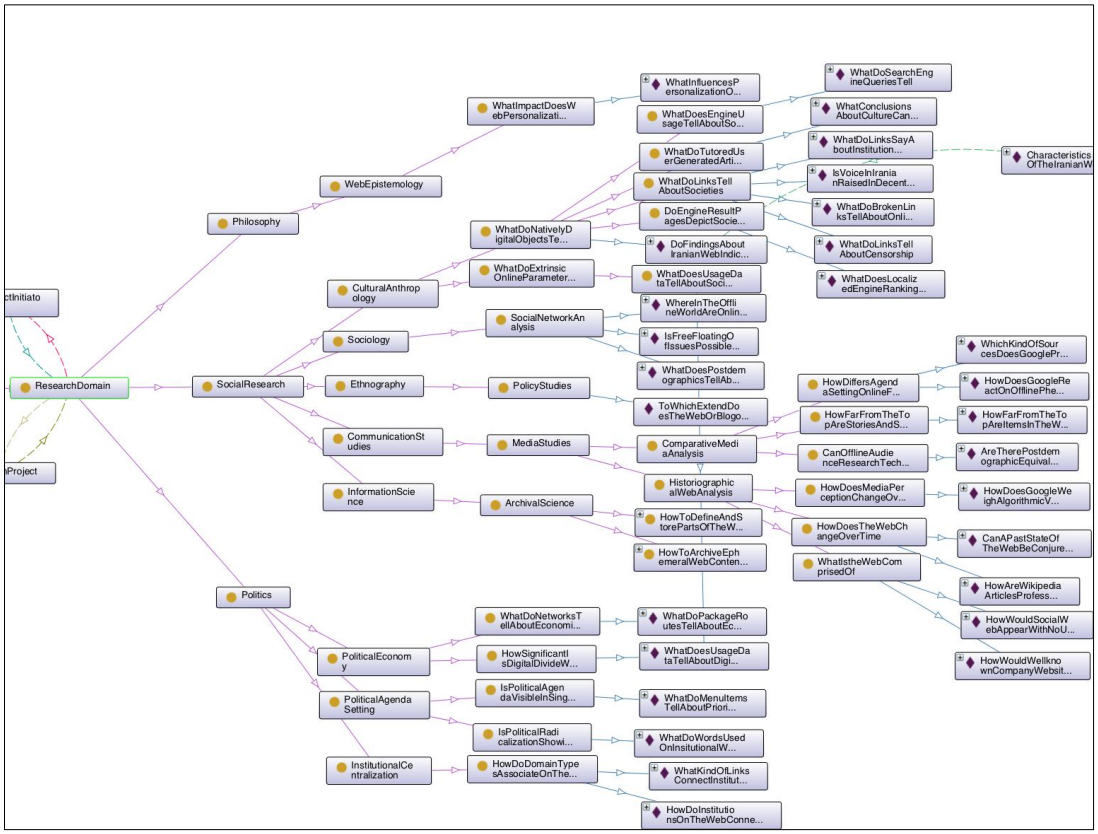

Illustration 5-5: The *ResearchDomain* Class Thread with all Contained Entities (Protégé export)

5.2 Prospective Use

The Digital Methods ontology resulted in a small text file. The non-profit nature of the Protégé editor and the underlying Web Ontology Language guarantee an independent use of the file outside of Protégé, and the possibility for manipulation in other applications as well as a reuse for multiple purposes in web technologies or others. Nevertheless, neither the file nor the editor per se supports sharing and interactive use (e.g. via a sharing functionality within the desktop application); a web-based collaboration is preferable. The Stanford University, originator of the Protégé editor, provides a web-based collaboration environment for any kind of ontology – supporting RDF/XML, Turtle, OWL/XML, OBO, and other formats – in a highly configurable user interface (Stanford University 2014b). It is a tool for allocation on a technical level as well as concerning the scientific audience, supporting professionally qualified exchange with other ontology creators. At current state, the Digital Methods ontology was uploaded, but hidden to the public. However, a future distribution into widely dispersed audiences via a simple hyperlink is conceivable.

6 Evaluation

»The conceptual point of departure is the recognition that the Internet is not only an object of study but also a source« (Rogers 2013: 23).

6.1 Introduction

The previous chapter described how the ontology evolved to be a taxonomical illustration of the Digital Methods. The fact that the ontology could be described without problems is in itself the first indicator for success, since errors would have led to invalid results in chapter 5. Additionally, some preliminary work was done to ensure correctness: the methodological and epistemological foundation (chapter 3) supported a well structured induction process, and the preliminary considerations on a suitable structure (chapter 4) standardized the procedure of decomposition and assembly in a new arrangement. However, there is obviously a strong need for a validation of both the process of induction and the results. The results again must be checked for both their validity in general and their possibilities of use. The following chapter hence illustrates three dimensions of evaluation:

1) *Result validity* means data cleansing is used to ensure that the ontology is logical in itself. The proposed validation method of »data cleansing« is grounded in content analysis and requires the results to be manually checked for logical errors and inconsistencies.
2) *Process reliability* refers to the fact that the induction process was aligned very closely to the object of study: Instead of applying some generic, well defined top-down framework, all items were induced in an ad hoc process along the Digital Methods domain described in the book. It is hence necessary to validate the generalizability of the ontology for the eventual integration of items that were *not* described by the very same author, but come from various other sources. The method applied is to gather a control group of research projects from various sources and evaluate whether they »fit« into the ontology.
3) *Utilization Quality* acknowledges that one of the most important objectives of the ontology is its future use by a professional audience. To evaluate whether the structure is comprehensive and the addition of new concepts is generally possible for future users, user stories have been established in chapter 2.3. These user stories will be continued in the following by adding some scenarios, in which the goals of the involved roles need to be fulfilled.

6.2 Result Validity

Traditional research methods like Content Analysis have stressed the error-proneness of empirical methods in which one or more researchers systematically describe media content. It acknowledges two metrics of reliability: *Intracoderreliability* and *Intercoderreliability*. The former depicts the consistence of one researcher during his coding work; the latter describes the homogeneity of two or more coders (Brosius, Haas & Koschel 2012: 151). Whereas Intercoderreliability is obviously not a possible constraint in the circumstances of this paper, the concept of Intracoderreliability is applicable to the current paper, because it is important to retrospectively identify falsifications that are likely to have occured within the process. Instead of following the interactive familiarization suggested by Content Analysis, the reasoner provided by Protegé, which was introduced in chapter 4.2.2, detects false occurrences. Additionally to technical error detection, the reasoner constantly checked the correct connections between toplevel classes, which have previously been assigned to *domains* and *ranges* (see Illustration 6-1). This concept allows for the reasoner to »know« that a research project must always:

1) Utilize some Digital Method,
2) Answer some research question from a specific research domain,
3) Be from a specific year
4) Have been conducted by a specific conductor.

Even without explicitly defining such a relation, the ontology will show it to any user once he starts the reasoner tool. That way, a user would for instance immediately see that the project *WhitehouseGovIssueList* was conducted *Utilizing* a method called *DocumentationOfSingleSiteHistory* – even if he had not seen that specifically, the kind of utilization was *AnalyzeAndCompareItemsInIssueListOnHomepage* (as a subordinate of *Utilizes)*. In the event that this user would not only browse the ontology, but planned to add something to it, he would immediately see that he is required to add a subproperty of *Utilize*, namely *AnalyzeAndCompareItemsInIssueListOnHomepage,* to specify the relation between the *WhitehouseGovIssueList* and *DocumentationOfSingleSiteHistory.* Concluding, as soon as one assigns *AnalyzeAndCompareItemsInIssueListOnHomepage* between two individuals, it is necessarily following that these two individuals are also connected via *Utilize*. This principle also served as a control mechanism by revealing false assignments, as Illustration 6-2 shows.

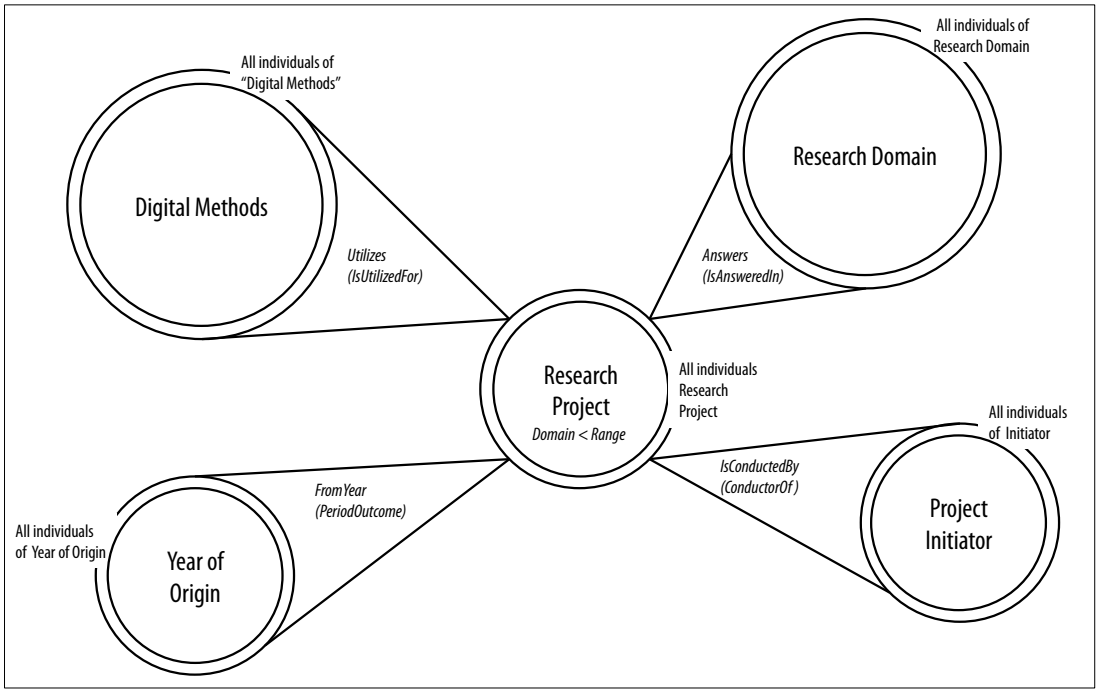

Illustration 6-1: Domains and Ranges of Toplevel Classes (own illustration)

As a second control mechanism, logical problems will be detected and solved manually. Whereas the reasoner can show falsely used statements in a formalized way, the following step of manual data cleansing will expose intellectual problems; for instance, since it was stated that any knowledge statement is initialized by a research project, and that every research project will have one related individual in the *ResearchDomain* and the *DigitalMethods* superclass, the total number of individuals in the three classes must be identical or the number of individuals in *DigitalMethods* or *ResearchDomain* must not exceed the number of individuals in *ResearchProject*. The classes *YearOfOrigin* and *Conductor* must be even with *ResearchProject*. Consequently, all orphan classes and individuals of were removed, and missing classes or individuals would have been added if necessary. Additionally, some orphan subclasses in *DigitalMethods* were removed, which had originally been created as tributes to important concepts that were described by Rogers, but not substantiated with applications in research projects. It was decided that since relationships among individuals create the value of the ontology, these orphan classes were disturbing.

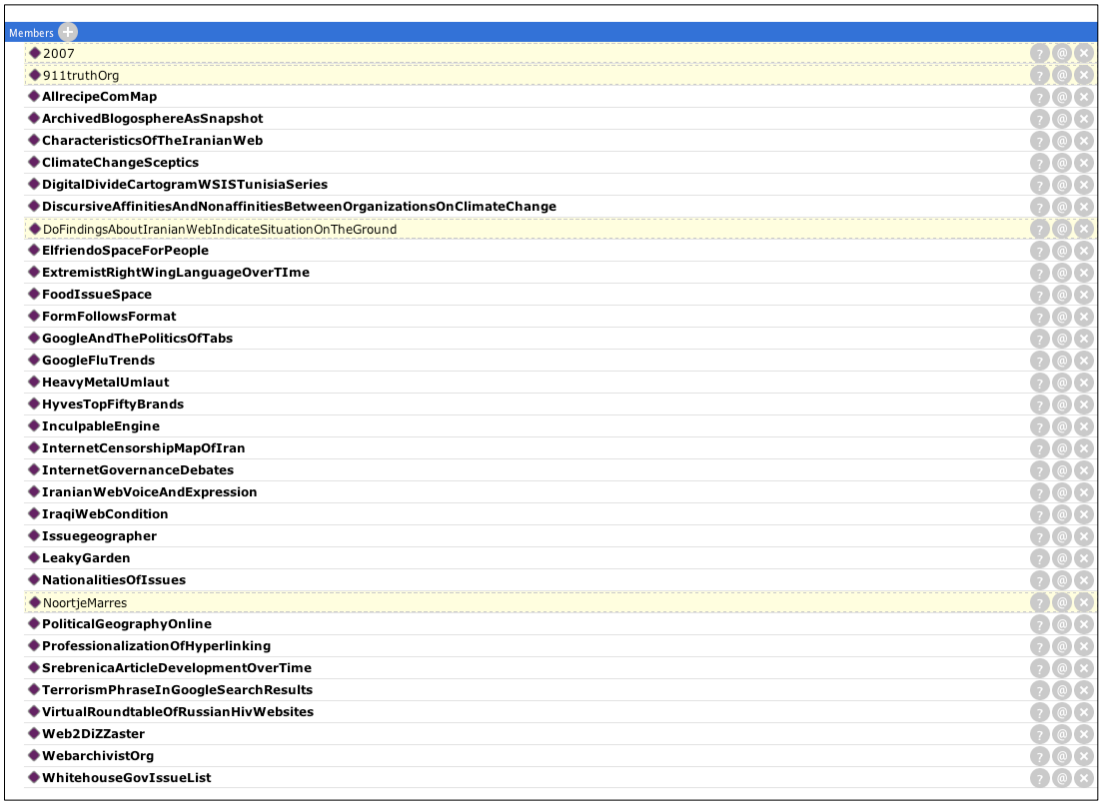

Illustration 6-2: False Attributions of Individuals to Class ResearchDomain (Protégé screenshot)

Secondly, classes can be checked for occurrences of false type individuals. For instance, members of the class *ResearchProject* can neither be in the form of a question nor a date; although no false individuals were directly associated with the *ResearchDomain* class, some attributions stem from falsely used statements of properties of individuals on lower levels, as the reasoner reveals (Illustration 6-2): yellow fields are only implicitly existent members of classes, which were detected and made explicit by the reasoner – in this case they show falsely used, alien elements: Dates, conductors, research questions have to be checked for their parameters and false attributions have to be eliminated. This is eased by the ability to navigate from individual to individual up to the culprit. In the example shown in the illustration, it was found that the individual *2007* was allocated falsely in a statement about an individual of a *ComparativeMediaAnalysis* subclass, namely *AreTherePostdemographicEquivalents-ToNielsenTvInterrogation,* as shown in Illustration 6-3.

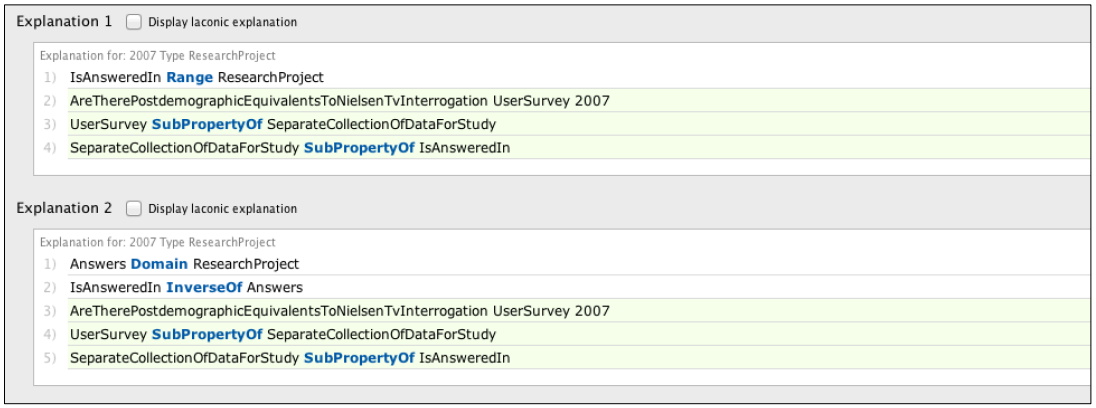

Illustration 6-3: Reasoner Explanation View (Protégé screenshot)

A closer look at the individual reveals that *AreTherePostdemographicEquivalents ToNielsenTvInterrogation* was allocated to *2007* with a subproperty of *IsAnsweredIn: UserSurvey* (Illustration 6-4).

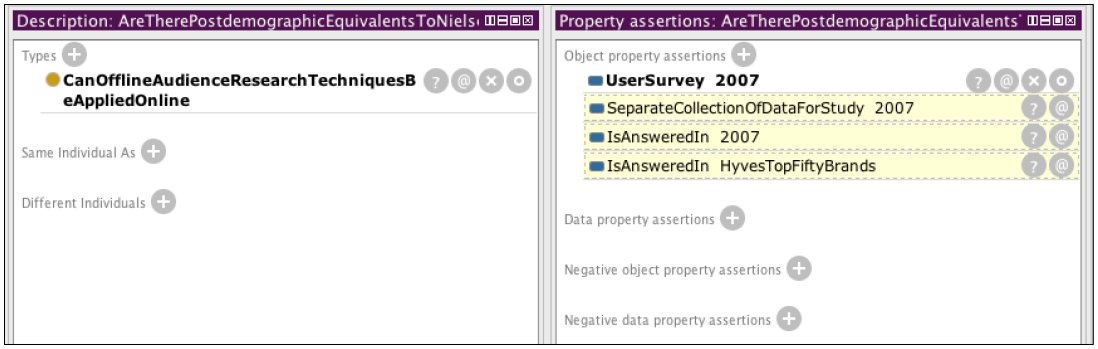

Illustration 6-4: Individual Object Property Assertions View (Protégé screenshot)

This was hence a matter of a careless mistake; the reasoner already indicates the appropriate attribution, which is the project *HyvesTopFiftyBrands.*

An equivalent technique resolves the falsely asserted individual *DoFindingsAboutIranianWebIndicateSituationOnTheGround* (Illustration 6-2); it was found that the individual was referring to itself instead of referring to the project *CharacteristicsOfTheIranianWeb. 911TruthOrg* however is a research project and was simply lacking the dedicated allocation – although the reasoner already and sufficiently shows the affiliation to the right class, it has been added manually for consistency reasons. Several inconsistencies were removed following this approach.

The next step of data cleansing concerns content-wise revisions and some disassociations from the origin. It occurred that once the ontology was detached from the book, some attributions of toplevel classes were not sufficient due to their non- self-explaining nature.

As for the *DigitalMethods* classes, some assumptions were adapted from Rogers, as explained already. This pertained the wording and distinction of certain subclasses and

individuals. However, this is limited to some areas, whereas it does not seem to fit for others. For instance, Rogers calls the three historiographical dimensions of web archiving *biographical*, *event-based* and *national.* This paper proposes a distinction into *event-based history of website*, *anatomy of website* and *historical site analysis.* This is more expressive and more clearly related to the different sizes of study objects (history of well-defined timespan vs. snapshot of current situation vs. longer time period), which is important in the ontology since there is no space for long explanations. Another dimension of this is the replacement of the expression »Conjuring up a past state of the web« (Rogers 2013: 77) to the favor of a subclass within the *LinkAnalysis* class, namely *HistoricalLinkAnalysis* – again, this is a more expressive term when stripped off context.

The *ResearchDomain* thread needs special subsequent attention, for it was constituted within the least rigid framing. The research questions themselves, included as individuals, and their allocation to certain research question classes on lower taxonomical levels, remain unaffected, since not the questions' specific relationships to Digital Methods and projects are questioned, but the distinctive and delimitative taxonomy of research domains that construe the thread.

Several adjustments were made in the *ResearchDomain* superclass:

1) The class OrganizationalPolitics, which held structural research on web governance and institutional concentration, was adjudged inappropriate because the term is used solely for internal organizational matters (Wikipedia 2014l).
2) During the cleansing process, it was considered to treat political science as a subordinate of social research, because the research projects related to this field show strong method-wise and subject-wise correlations to social sciences, and because it is regarded as belonging to social science in some occurrences, e.g. in Wikipedia (2014j). However, Wikipedia (2014h) defines political science as a generic field, which is why it appears legitimate to remain a sibling class independent from social science.
3) Research concerning governmental censorship should obviously be allocated to Politics. However, the similarity of methods or contentual vicinity demands for the inherent individuals to remain in the subclass of *CulturalAnthropology,* in which also their (non-political) siblings and counterparts are. This reveals a logical problem of the simplification of statements within a domain into small pieces of information with bidirectional relationships: Ambiguity of items is not envisaged, yet occasionally required. At current state, the context-driven allocation into one class and dismissal of the other appears the best solution, yet for prospective scalability, it is not sufficient.

4) The class *ArchivalScience*, subsuming the challenges of archiving the web and its ephemeral content, was moved from communication studies into the distinct class of *InformationScience*: Although information science makes use of methodology from informatics and diverse branches of social research – among others communication science – it is not directly affiliated with a certain branch (Wikipedia 2014d).
5) *NetworkTheory* was renamed in *SocialNetworkAnalysis* to make it distinct from computer science, and removed to the independent class sociology, the branch were this method origins (Wikipedia 2014i).
6) Ethnomethodology has some relation to several branches, methods or projects within the ontology. However, it was decided to exclude this domain from the ontology due to the ambiguity of definitions that apparently exist:

   »One of the most perplexing problems for those new to ethnomethodology is the discovery that it lacks both a formally stated theory and a formal methodology. As serious as these problems might appear on the face of it, neither has prevented ethnomethodologists from doing ethnomethodological studies, and generating a substantial literature of 'findings' « (Wikipedia 2014c).

   On a metalevel, Wikipedia stresses that its users found no consensus about the reliability or accuracy of the article, which adds to the impression of a deficient definition of the field.
7) The superclass *BigDataAnalysis* was removed; despite its importance for social research, big data analysis had no equivalence in any research project and was hence orphaned.
8) Furthermore, some orphan properties were removed that had no references to any individual.

Apart from this complex and important *ResearchDomain* thread, some other adjustments have been made. The class *ConductConcerns,* a subordinate of *Concerns*, has been renamed to *ConceptConcerns* because the former name was perceived ambiguous: As opposed to its counterpart sibling class, *MethodologicalConcerns,* it shall precisely *not* point at methodological concerns about the way it was conducted, but rather at general criticism that is existent about its conductor, its content or related concepts. An example is Google as a research tool about societal knowledge, as in the example of *GoogleFluTrends*; a general criticism, which is not directly related to that one specific study, is the perception of Google as a G*atekeeper* of information.

6.3 Process Reliability

The second evaluation method concerns the process of creating the ontology. One of the general weak spots of designing ontologies, and hence of the method applied in this paper, is its flexible logic and the consequential dependency on rather arbitrary seeming heuristics for decisions: Questions like »What knowledge is included, what not?«, »On which abstraction level shall an individual reside?«, »To what other individual is it connected, and by which property?« etc. are not answered in a generic scheme, but have to be applied closely to the knowledge domain: »There is no one correct way to model a domain – there are always viable alternatives. The best solution almost always depends on the application that you have in mind and the extensions that you anticipate« (Noy & Mcguinness 2000: 4). The interchangeable structure of items is hence on one hand a guarantee for a suitable representation of any domain, but on the other hand it misses a guiding structure, and insofar lacks a control instrument. Additionally, the Digital Methods ontology so far is solely based on research introduced by *one* single author. Concerning future scalability, it has to be evaluated whether descriptions of other researchers fit into the contentual approach of Rogers and the structural approach of this paper.

The solution to this is to gather another set of studies that have not been described in the book, and check whether this control group fits into the ontology. Due to the illustrated flexibility, this should work per se with any other item. However, having in mind the desired inverse tree structure of the present ontology, all new studies have to fit into the general structure that already exist: A (digital) method, a related research domain, a conductor and a year of origin should always be applicable; further information should be integrable into the additional classes *Concerns* and *Concepts.*

As a result, the process described in Illustration 4-4 needs to be applied to this new set of studies. Five research projects were retrieved and translated into ontology items. To ensure their randomness in order to be significant, a diversity of sources was used; studies have been retrieved in the ACM Digital Library, the IEEE Digital Library, EBSCO Host, and other relevant databases, as well as in the Cologne University of Applied Sciences eBook library and the visible web (Google).

#Ausvotes: Twitter Activity Patterns Across Electorates
Track Twitter activity patterns (*Tweets* and *Mentions*) around Australian federal politicians' and candidates' electorates in a map (Bruns 2013).

a) Individual *AusvotesTwitterActivityAcrossElectorates* in class ResearchProject
b) Individual *TwitterActivityPatterns* in class *IssueAnalysis* (new subclass of *SocialMediaResearch* → *DigitalMethods*)
c) a) is connected to b) via property *VisualizeTwitterActivityPatternsAboutFederalPoliticiansAndCandidates* (subproperty of Utilize)
d) Individual *DoLocalElectoralRacesShowUpOnTwitter* in class *IsInfluenceOfOfflineIncidentsOnUserGeneratedContentVerifiable* (new subclass of *MediaStudies* → *CommunicationScience* → *SocialResearch* → *ResearchDomain*)
e) a) is connected to d) via property *ApplyColorsToMapOnALogarithmicalScale* (subproperty of *Answers*)
f) Individual *2013* in *YearOfOrigin* (federal election campaign)
g) Individual *MappingOnlinePublics* in the classes *EducationalScientificBackground* and Institution

The complete integration of this research project is possible by adding two subclasses (in *ResearchDomain* and *DigitalMethods*).

Table 6-1: Control Group Object 1 – #Ausvotes: Twitter Activity Patterns Across Electorates

Social Media as a Measurement Tool Of Depression in Populations
Feasibility study of leveraging social media postings to understand depression in populations (De Choudhury, Counts & Horwitz, 2013).

a) Individual *SocialMediaAsMeasurementToolOfDepression* in class ResearchProject
b) Individual *SocialMediaDepressionIndex* in class *SentimentAnalysis* (new subclass of *SocialMediaResearch* → *DigitalMethods*)
c) a) is connected to b) via property *GatherDataAndDeriveTrainedCorpusToDeriveMetricsForIndex* (subproperty of *Utilize*)
d) Individual *WhatDoesMicrobloggingActivityTellAboutSociety* in class *WhatDoNativelyDigitalObjectsTellAboutSocieties* (subclass of *CulturalAnthropology* → *SocialResearch* → *ResearchDomain*)
e) a) is connected to d) via property *RevealGeographicalDemographicSeasonalPatternsOfDepression* (subproperty of *Answers*)
f) Individual 2013 in *YearOfOrigin*
g) Individual *MicrosoftResearch* in the classes *CommercialBackground* and Institution

The complete integration of this research project is possible by adding one subclass (in *DigitalMethods).*

Table 6-2: Control Group Object 2 – Social Media as a Measurement Tool of Depression in Populations

Traditional Media Seen from Social Media

Analyse Twitter for insights into media supply and demand landscape (An et al., 2013)

a) Individual *TraditionalMediaSeenFromSocialMedia* in class ResearchProject
b) Individual *TwitterActivityPatterns* in class *PostdemographicsForAdvertisersResearch* (subclass of *SocialMediaResearch* → *DigitalMethods*)
c) a) is connected to b) via property *AnalyseTwitterSubscriptionAndInteractionForInsightsIntoMediaLandscape* (subproperty of Utilize)
d) Individual *WhatDoesTwitterMentionAndSubscriptionTellAboutMediaLandscape* in class *WhatDoesMicrobloggingActivityTellAboutSociety* (new subclass of *WhatDoNativelyDigitalObjectsTellAboutSociety* → *CulturalAnthropology* and *MediaStudies* → *CommunicationStudies* → *SocialResearch* → *ResearchDomain*)
e) a) is connected to d) via property *RevealMediaSupplyAndDemandLandscapesThroughEvaluatingInterpersonalNetworksAndStoryPropagation* (subproperty of Answers)
f) Individual 2013 in *YearOfOrigin*
g) Individuals *JisunAn*, *DanieleQuercia*, *MeeyoungCha*, *KrishnaGoummadi*, *JonCrowcroft*, connected by property *FormsResearchTeamWith* in the classes *EducationalScientificBackground* and *SinglePerson*

The complete integration of this research project is possible by adding one subclass (in *ResearchDomain).*

Table 6-3: Control Group Object 3 – Traditional Media Seen from Social Media

The Geographically Uneven Coverage of Wikipedia

Discover biases of Wikipedia's articles in their geographic distribution (Oxford Internet Institute 2012).

a) Individual *GeographicallyUnevenCoverageOfWikipedia* in class *ResearchProject*
b) Individual *CrosscountryComparisonOfLocationReferencesInWikipediaArticles* in class *WikipediaAsCulturalReference* (subclass of *DigitalMethods*)
c) a) is connected to b) via property *AnalyseMentionsOfPlacesEventsAndPeopleThroughoutWikipediaLanguageVersions* (subproperty of Utilize)
d) Individual *WhatConclusionsAboutLocationDominanceCanBeDrawnFromWikipediaArticles* in class *WhatDoGeotaggedUserGeneratedArticlesTellAboutDominanceOfCountriesInKnowledgeRepositories* (new subclass of *WhatDoNativelyDigitalObjectsTellAboutSociety* → *CulturalAnthropology* → *SocialResearch* → *ResearchDomain*)
e) a) is connected to d) via property *CorrelateGeotaggedArticlesWithWorldMap* (subproperty of *Answers*)
f) Individual 2012 in *YearOfOrigin*
g) Individual *OxfordInternetInstitute* in the classes *EducationalScientificBackground* and *Institution*
h) Individual *BiasesOfWordCountAnalysisTroughLinguisticDensityOrVerbosity* in the class *CrosscomparisonOfNationalArticles* (subclass of *MethodologicalConcerns* → *Concerns*)

The complete integration of this research project is possible by adding two new subclasses (in *ResearchDomain* and *Concerns*).

Table 6-4: Control Group Object 4 – The Geographically Uneven Coverage of Wikipedia

Top 10 Twitter Languages in London Detect languages of Tweets in the London area (Lima 2014a). a) Individual *TwitterLanguagesInLondon* in class ResearchProject b) Individual *MapLanguagesOfTwitterTweetsInGeographicalArea* in class *PostdemographicsForCulturalResearch* (new subclass of *DigitalMethods* → *SocialMediaResearch*) c) a) is connected to b) via property *Analyse3MillionTweetsForLanguageAndCreateColorCodedGeographicalMap* d) Individual *WhatConclusionsAboutMulticulturalSocietyInUrbanAreaCanBeDrawnFromTweetLanguages* in class *WhatDoesMicrobloggingLanguageTellAboutSociety* (new subclass of *WhatDoNativelyDigitalObjectsTellAboutSociety* → *CulturalAnthropology* → *SocialResearch* → *ResearchDomain*) e) a) is connected to d) via property *AlgorithmicCollectionOfLanguagAndGeolocationOfTweets* (subproperty of *Answers*) f) Individual 2012 in *YearOfOrigin* g) Individuals *EdManley* and *JamesCheshire*, connected by property *FormsResearchTeamWith*, in the classes *EducationalScientificBackground* and *SinglePerson* The complete integration of this research project is possible by adding two new subclasses (in *DigitalMethods* and *ResearchDomain*).

Table 6-5: Control Group Object 5 – Top 10 Twitter Languages in London

More research projects were found, some of which fulfil the definition of Digital Methods only at first glance. An example is Debin et al. (2013), who conducted a web-based Delphi survey proposed to 288 influenza experts to determine the accuracy of previously determined influence epidemic data based on statistical models. The experts were invited to draw starting and ending weeks of influence epidemics in France from 1985 and 2011 in 32 time-series graphs, grounded on the previously gathered statistical offline data. Here, the web was solely used as an instrument of surveying, which is why the research project, although interesting, was not used in the control group.

Of all retrieved control studies, none provided any difficulties for the existing ontology. It has to be stated, though, that the control group consists only of new studies, not of entirely new *concepts*. Whereas this is sufficient to assay the process reliability as desired, concepts would require the ontology to change more significantly, as the following example shows: A term that Lev Manovich refers to is *Cultural Analytics*, including studies like »Wikipedia Edits during the Middle-East Protests« by Elijah Meeks in 2011 (Lima 2014b), in which a short, dynamic visualisation of Wikipedia article changes made the types of edits accessible with color-coding. Another application of Cultural Analytics is »Making Visible the Invisible« by George Legrady, in which library transactions (lending of media such as books, DVDs, CDs) were illustrated in real-time on six large LCD screens in the foyer of the Seattle Central Library during the years 2005 – 2014 (Legrady 2014). These approaches to visualising large data sets is an important concept of the Digital Methods domain, which is why it would have to be integrated in one of the superior classes within *DigitalMethods*.

6.4 Utilization Quality

The need for a user based evaluation has already been illustrated in chapter 2: Two essential use cases were identified for ontology engineers as well as users: either putting something into the ontology, or taking something out of the ontology (see Illustration 2-1).

One already introduced anomaly of this user evaluation is that the ontology is not a concrete system, and the »mechanical« interaction of a user with some desktop software is *not* part of the evaluation. Instead of finding out whether someone would be able to interact with some software that processes OWL (Protégé or any other tool), it is much more important for him to understand what the classes, individuals and properties in this present ontology *represent,* therefore what information they provide to understand the ontology as a whole. This is even more important since ontologies may serve as meta-models that other applications, e.g. web based systems, are based upon – of which the interaction is absolutely unpredictable at this point. Hence, the following user stories and related scenarios put a focus on the abstract concepts of understanding, exploring and learning, instead of testing system-wise actions and reactions like clicking or opening.

It is further assumed that any user role in the following scenarios has already decided whether he would prefer exploring the ontology with help of a visual tool like OWLViz for Protégé, which allows for »class hierarchies in an OWL Ontology to be viewed and incrementally navigated« (Protégé Wiki 2013), or prefer to see all items in lists and divided into classes, individuals and properties, as the default view of Protégé suggests – or any other possible scheme. Having stated this preliminary, the user stories and scenarios will formulate actions in the sense of "she navigates through the classes", which shall cover all possible appearances; what matters for the scenarios is the process of construing knowledge and adding information into the right location, which is (more or less) detached from the visual appearance. The following section will repeat the user stories from chapter 2, provide two scenarios for each, the first one being »desirable« and the latter »alternative«, and will afterwards illustrate the process of construing the necessary knowledge with a snapshot of the final state, therefore the result described in the scenarios. Whereas user stories 1 to 3 are assumed to be best illustrated with the network view provided by OntoGraf, user story 4 concerns a fictional ontology expert and is hence illustrated with a view of the complete XML based syntax (Illustration 6-8).

User Story 1: Find research projects that concern search engine usage and its impact on societies, and evaluate the related methods for their ability to be reused for own research project about the political landscape within a language sphere manifesting in search phrases.

Desired Scenario: The political scientist understands that what she sees first are the high-level concepts of the knowledge domain, which serve as some sort of overall classification of what the ontology contains. She interprets that, for her purpose, of interest are primarily the classes *DigitalMethods*, *ResearchDomain* and *ResearchProject* – whether or not she understands that they are called classes is not important, since the hierarchical structure indicates the difference between classes and individuals. She understands that she can navigate through the branches in the form of a path, where she has control over all decisions about junctions. By randomly exploring the ontology superclasses, she reaches *SearchQueryAnalysis* within the *DigitalMethods* class and is able to see all research projects that have been collected in this thread. She stumbles upon *AllRecipeComMap*, a study about creating geographical maps of US population's recipe preferences, and she discovers that someone added localization to search data just as she plans to do. She decides that this study is relevant for her purposes and that she has to further inform herself about it; with help of the provided information about subject, authors and year of origin, she is able to do a web research and retrieve a report. The ontology also provides her with the location of the research project description in the Digital Methods book (Rogers 2013: 5). Her needs are satisfied.

Alternative Scenario: The political scientist does not understand the superstructure. She starts exploring the high-level concepts of the ontology without knowing that *DigitalMethods* is a synonym for *method* or *methodology*. Since the high-level options are few, she starts exploring the research projects. At one point, she finds that the descriptive title of one project strongly indicates search engine usage as the underlying data set, and starts exploring its neighbourhood. All attributes of this project become visible, among others the method that was deployed. From this method, she is able to retrieve related projects and similar methods. She also identifies the more abstract method group and is now able to find all projects processing search engine data. She finds that *AllRecipeComMap* is the most relevant project for her purposes, and that she has to further inform herself about it.

Table 6-6: Scenarios for User Story 1 – Retrieve Methods Suitable for Reuse

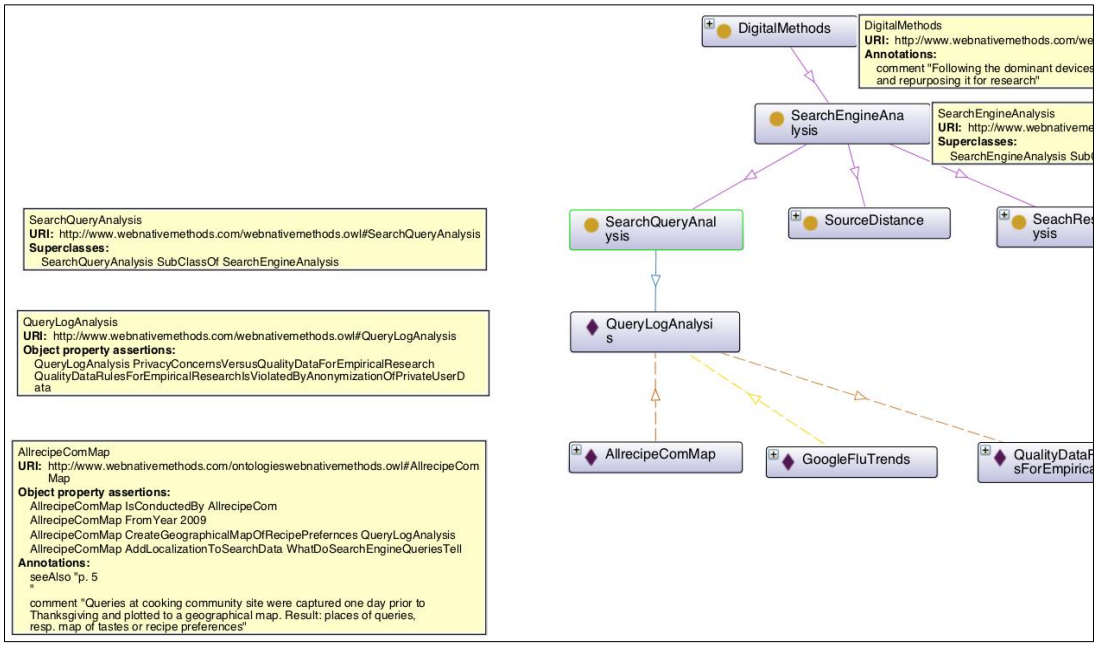

Illustration 6-5: Exemplary User Journey for User Story 1 – Desired Scenario (Protégé export)

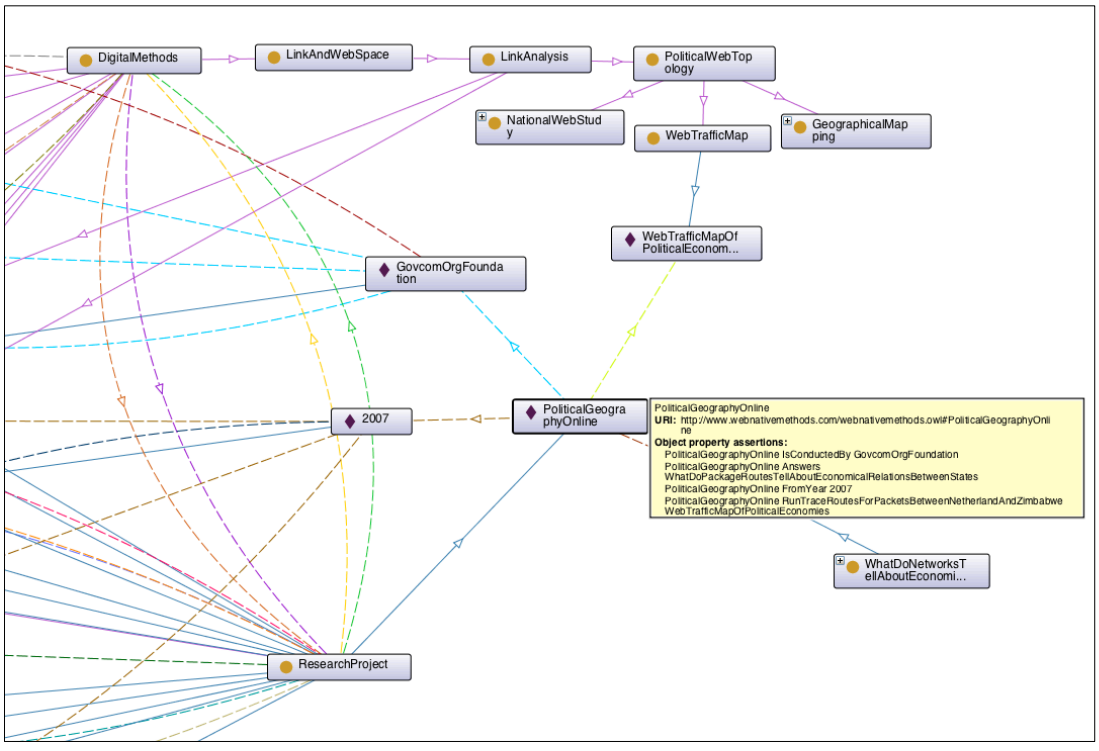

Illustration 6-6: Exemplary User Journey Around »Political Geography Online« for User Story 2 – Desired Scenario (Protégé export)

User Story 2: Explore the ontology to learn about the scientific domains that have been researched with web-native methods so far, and integrate own results of a research project in the domain of cultural studies that was conducted based on ratings on the Internet Movie Database (IMDb).

Desired Scenario: The cultural scientist has recently closed a research project about Automated Sentiment Analysis (Data Mining) of movie ratings in the Internet Movie Database (IMDb). After someone tells him about the Digital Methods ontology, he plans to expand it with his own findings, which he identifies as highly relevant for this field. He explores the described projects in the *ResearchProjects* class to see whether his contribution was a repetition of already existing context. Since he cannot find any projects that carry the IMDb in their names, or any other hints on similar projects, he decides to add an individual in that class. He hesitates as he discovers that there is no further information assignable except for the name of the studies. He aborts and goes back to exploring, quickly finding that the essential information about every study is contained in the properties that link them to other individuals. Exemplary, he follows the relationships of the project *PoliticalGeographyOnline* to discover and understand that it is related to four other concepts. Whereas conductor and year of origin are self-descriptive, the relation to *DigitalMethods* is not. The researcher follows the path up from the individual connected to his project (*WebTrafficMapOfPoliticalEconomy*) to learn that he can see the applied method here (a subordinate of *LinkAnalysis*, as shown when reaching the higher levels of the branch). This motivated him to go back and explore the other related individual and its branch, namely the question *WhatDo-NetworksTellAboutSociety*. Afterwards, he is able to apply the general structure to his own intend and derive the necessary properties and individuals of other classes. His contribution to the knowledge domain in the end consists of the description of a research project, its conductors and year of origin, the methodological approach and a research question in the class ResearchDomain. Since he did not know where exactly to put it in the scientific taxonomy, he opened a new superclass called *CulturalTheory*, and put his question inside in the form of an individual.

Alternative Scenario: The cultural scientist explores the ontology as described above, and decides that his research is a valuable contribution. He plans to put a new entry in the *ResearchProject* class, but hesitates as he discovers that there is no further information assignable except for the name of the studies. He continues to add his study and finds a window called *Annotations*. He copies a short abstract from his publication into the Annotations window, providing all information about the study that he finds valuable in this context, and closed the dialog. Subsequently, some other researcher explores the ontology. It is not his first visit, in fact, he has been using this web-based ontology for several months now and it has been very helpful. He discovers a new entry, and further inspection shows him that all information is »hidden« within the Annotations, which are not machine-readable, as he knows. He decides to move the knowledge from there into the classes by splitting it into individuals and properties, until the research project about IMDb is a legitimate part of the ontology.

Table 6-7: Scenarios for User Story 2 – Integrate Own Findings in Appropriate Location

User Story 3: Scan the ontology for all conductors of studies to find own name, and from own name follow outgoing paths to other information, such as studies by this author, methods used in these studies, questions asked in these studies, motivation to conduct these studies, etc.

Desired Scenario: The researcher intends to find his own name and the concepts that have been assigned with it to prove whether they conform to his initial ideas. After seeing the toplevel classes, he immediately understands that he does not need to find his project or method without knowing the assigned names, but can quickly search for his name instead by exploring the class *ResearchConductor*. He discovers the name of his research partner as an individual of the class and understands that it provides further information about what he did by browsing the related concepts. One of the properties assigned to his research partner, *BruceEtling,* is *FormsResearchTeamWith.* Following this path, he discovers his own name and sees a connection to *IranianWebVoiceAndExpression,* the study he has conducted. Via concrete descriptions of its relation to other concepts, such as a research question (individual) and how it was answered (property), the big picture of his research reveals. He stumbles upon the related method's name (*NationalWebHealthAnalysis*), which he himself had never heard of, and counterchecks the Digital Methods book (Rogers 2013) to see where the term originates. Noticing that it is used in the book to classify several studies with a similar topic, he turns to the ontology again and starts exploring this field. He discovers projects of other researchers that apply a similar method, such as a project about the Iraqi web condition, and decides to contact the conductors previous to his next research project.

Alternative Scenario: The researcher discovers his name after the process described above, but instead of being content with the properties of the research project as they are displayed within the properties and individuals, he discovers false statements about the way the study was conducted. He counterchecks the book that the ontology refers to in its description (Rogers 2013) and finds that here, all attributes are correct. Since the ontology is online to the public, the researcher desires the statements to be fixed. He turns back to the ontology to search for a contact person. In the help section, he instead discovers that he is able to rework the ontology by himself. After a short familiarisation, he knows how to adjust the concerned parts, and corrects everything.

Table 6-8: Scenarios for User Story 3 – Retrieve Information about own Project and Evaluate Correctness

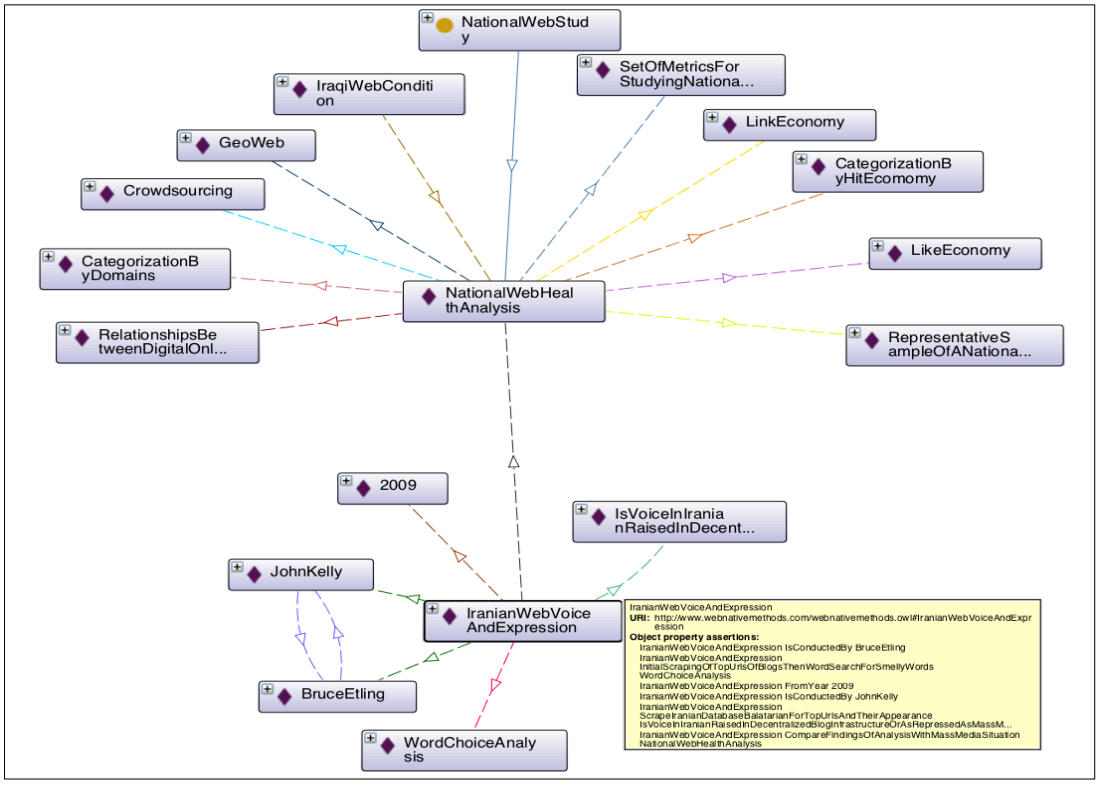

Illustration 6-7: Exemplary User Journey for User Story 3 – Desired Scenario (Protégé export)

User Story 4: Explore the ontology and comprehend the logic upon which it builds, estimate its significance for the field of web science and reuse it entirely or partly to place it in a broader context. Desired Scenario: The researcher from the field of web science is conducting a literature-review based prevalence study of research about the web. The initial exploring phase shall result in a prospective development of a classification scheme for all research related to web science. She stumbles upon the Digital Methods ontology and finds that this display format is well suitable to describe a research field in that domain. She starts exploring it with the desire to understand the underlying structure of thinking as well as the degree of formalization. She discovers that the ontology contains no definitions of data property assertions, but is in itself consistent and correct from a technical point of view, and since she already considered presenting her literature research results in a knowledge representation, which she finds best suitable for her meta-study, she decides to construe a broader subject ontology with the Digital Methods ontology being one branch of it. The logical structure is adjusted throughout this process, but the general approach to Digital Methods, including its allocation into the three high-level concepts of method, project and scientific domain, remains. Alternative scenario: The researcher explores the field of web science and discovers the ontology as described above, but she does not intend to display her own research results in a similar form. Yet, she perceives the ontology as valuable to use it as a starting point for her research in this web science subsection. By exploring threads and deeper investigating single concepts, over time she thoroughly understands the Digital Methods definition and what it consists of. She is able to demarcate it from other web research, such as those that perceive and analyse the web as a medium, and becomes capable of retrieving more applications of web-native methods in a new set of studies from several online repositories.

Table 6-9: Scenarios for User Story 4 – Comprehend Knowledge Domain and Reuse it for Broader Context

```
    <!--
    ///////////////////////////////////////////////////////////////////////////////////////
    //
    // Individuals
    //
    ///////////////////////////////////////////////////////////////////////////////////////
     -->



    <!-- http://www.webnativemethods.com/ontologieswebnativemethods.owl#AllrecipeComMap -->

    <owl:NamedIndividual rdf:about="&ontologieswebnativemethods;AllrecipeComMap">
        <rdf:type rdf:resource="&webnativemethods;ResearchProject"/>
        <rdfs:comment>Queries at cooking community site were captured one day prior to Thanksgiving and plotted to a geographical
map. Result: places of queries, resp. map of tastes or recipe preferences</rdfs:comment>
        <rdfs:seeAlso>p. 5
</rdfs:seeAlso>
        <webnativemethods:FromYear rdf:resource="&webnativemethods;2009"/>
        <webnativemethods:IsConductedBy rdf:resource="&webnativemethods;AllrecipeCom"/>
        <webnativemethods:Utilizes rdf:resource="&webnativemethods;QueryLogAnalysis"/>
        <webnativemethods:CreateGeographicalMapOfRecipePrefernces rdf:resource="&webnativemethods;QueryLogAnalysis"/>
        <webnativemethods:Answers rdf:resource="&webnativemethods;WhatDoSearchEngineQueriesTell"/>
        <webnativemethods:AddLocalizationToSearchData rdf:resource="&webnativemethods;WhatDoSearchEngineQueriesTell"/>
    </owl:NamedIndividual>



    <!-- http://www.webnativemethods.com/ontologieswebnativemethods.owl#GoogleFluTrends -->

    <owl:NamedIndividual rdf:about="&ontologieswebnativemethods;GoogleFluTrends">
        <rdf:type rdf:resource="&webnativemethods;ResearchProject"/>
        <webnativemethods:FromYear rdf:resource="&webnativemethods;2007"/>
        <webnativemethods:IsConductedBy rdf:resource="&webnativemethods;Google.org"/>
        <webnativemethods:Utilizes rdf:resource="&webnativemethods;QueryLogAnalysis"/>
        <webnativemethods:CompareSearchWithOfflineMedicalData rdf:resource="&webnativemethods;QueryLogAnalysis"/>
        <webnativemethods:Answers rdf:resource="&webnativemethods;WhatDoSearchEngineQueriesTell"/>
        <webnativemethods:GroundSearchQueriesInOfflineMedicalData rdf:resource="&webnativemethods;WhatDoSearchEngineQueriesTell"/>
    </owl:NamedIndividual>
```

Illustration 6-8: Exemplary Extract of the Ontology's XML Export as Potentially Used in User Story 4 – Desired Scenario (Protégé export)

6.5 Conclusion of Evaluation

The previous chapters have shown how all problems could be solved in the three evaluation dimensions. The result validity checking was important for detecting false statements, which would have led to errors. However, the error detection was not limited to technically correct or incorrect. Rather, the described manual cleansing contributed to a significant improvement of the ontology's meaningfulness. Next, the reliability of the results, which was checked with help of a control group, showed that the ontology is suitable for various projects from the domain of Digital Methods, independently from each author's own perception of this term: Any average research documentation provides enough basic information to be decomposed and assembled as ontology items. It also showed that when completing meta studies like this, it is generally advisable to retrieve work from various sources instead of grounding the collection and classification solely on one author or research initiative. It was for instance important to find that a lot of research currently focuses on social media analysis or big data. This is an expectable outcome considering the current discourse about it in various scientific domains, but it is a significant difference to the collection based on Rogers, in which social media and big data are mentioned, but in a rather abstract way. Finally, the user stories and scenarios tried to anticipate the main usage motivations and proposed solutions to these situations of use. This part is quite

challenging: Although there were no problems identified within the scenarios and they were each perceived as successful, it might be problematic to perceive the user related evaluation as completely successful. The reason for that has been insinuated in chapter 2: not a system is tested, but an ontology. Apart from the higher abstraction level, this also means that the software, with which a user interacts, is *not* part of any scenario. Of importance is *only* the more abstract concept that manifests as lists, graphs or any other form that the software in use provides. Nevertheless, having previously stated the desire to test the developed user stories based on scenarios, this is perceived as accomplished.

All in all, since no major problems occurred in the previous sections, the evaluation of all three dimensions showed that the ontology can be perceived as valid, reliable and usable.

7 Interpretation

»The issue no longer is how much of society and culture is online, but rather how to diagnose cultural change and societal conditions by means of the internet« (Rogers 2013: 21).

7.1 Introduction

It was said in chapter 3.3.1 that research in empirical social science strives for a generalization of observations to make statements about a social context. The equivalent of social context in the regard of this paper is, of course, the field of social and humane research with web-native data. If generalizable results could be achieved in the ontology, it would now be possible to deduce some statements about the research field of Digital Methods and its anchorage in traditional scientific domains, as well as findings about research conductors, their motivations and the circumstances in which they worked, as well as interest in certain methods or answers over time. This chapter is hence a first attempt of *interpretation* of the previously summarized results.

The ontology in its current form provides several starting points for this interpretation: From all the previously introduced superclasses, subclasses, individuals and properties, one can derive some assumptions and deduce hypotheses about the concepts that they describe. This would result in isolated considerations of all ideas, e.g. in the form of »All in all, research projects that use Digital Methods are rather short. Maybe this is due to the smaller trust in online data compared to offline data, or maybe due to the smaller number of researchers that are familiar with web-native methods«. Alternatively, one could pursue a cross-comparison of two concepts within the ontology, e.g. projects and time spans: »Research projects that use Digital Methods were rather short in the beginning, but it seems like four years ago, most of the research periods were extended considerably. At the same time, the interest in researching online social networks increased, as the *ResearchDomain* class shows«. A third approach might be to compare ontology spaces with certain external information, such as: »Research projects that use Digital Methods were rather short in the beginning, but it seems like four years ago, most of the research periods in German language projects were extended considerably. It is possible that this is related to the launch of Facebook in Germany«.

It has to be stated, though, that at current state, quantified statements about the research field – such as »35% of all researcher use search engine data for analysis« – do not appear legitimate, since they would require a bigger sample size to be significant.

Instead, *qualitative* statements about specific aspects can be deduced, which means individual occurrences are interpreted in an open, explorative process.

7.2 The Current State of the Digital Methods Research Field

Although an exhaustive scan of Rogers' work resulted in capturing the entire pool of Digital Methods research projects introduced in his work, the field of research that can be subsumed under the *Digital Methods* term is obviously not *limited* to his work (which became already apparent in the process reliability evaluation with a new set). Fortunately, if the ontology at hand is generalizable as desired, it is possible to derive general statements about the *entire* Digital Methods research field, although it is only partly and not thoroughly illustrated in the ontology. Consequently, the following first attempt of interpretation will make statements about the research field based on the nature of the toplevel classes, the contained individuals and their relationships (properties). Initially, one would possibly like to know how Digital Methods and research intentions of traditional sciences correlate in general. For this desire, not the classes and contained individuals are most relevant, but the property that connects *DigitalMethods* and *ResearchDomain*: The thread of the property *Answers,* the collection of all interest of various scientific domains in research projects, provides crucial insights. Based on this collection, six general areas of interest of web-native research, manifesting in research questions, are identified that differ significantly in terms of dimension, attributes and epistemological interest:

- Creation and evolvement of networks (link associations, package routes, circulation of information)
- Localization of offline phenomena in the online (search data based predictions, online vs. offline discussion occurrences)
- Censorship maps (demarcate national web(s), discover blocked network nodes and disconnected locations)
- Content (word choice development, language development, article tenor over time)
- Visual appearance (frames of sites without content, presentation of individual information)
- Issue context (contextual development of issues in the whole web, search engines promoting offline media subjects, debate culture on deep pages)
- Gatekeeping of search engines (results in different cultures as mutual agreement on issues, emphasis of sources over others, source removal)

In the fundamental work about the web of Berners-Lee et al. (2006a: 71), the authors identify largely correlating areas of interest of social aspects on the web:

- Web epistemology
- Web sociology
    - Communities of interest
    - Information structures and social structures
    - Significance and its metrics
    - Trust and reputation
    - Web morality and conventional aspects of web use

It shows that some spaces are congruent, although called slightly different, e.g. social and information structures, significance and its metrics, communities of interest. It also shows that with trust and web morality, some fields that Berners-Lee identify as crucial for web science, have not been tackled by the Digital Methods research so far. This is a first answer to the question whether one could identify correlations of methods or methodological gaps, which was raised in the introduction.

When looking at the Answer property's inverse, *IsAnsweredIn,* assumptions on the foundation of research data, like their origin in offline and online, can be derived:

- Two contrary strategies of gaining data are present: the repurpose of existing data, and the separate, dedicated collection of data for one research project.
- Within the field of research that repurposes existing data, e.g. the analysis of search engine queries over time, data is either solely web-native, or web-native but grounded in or combined with offline data.
- By far the most frequent form of collection when repurposing existing data is to subsequently gather the data after it is formed through usage; a concurrent collection of data, e.g. logging search engine queries in real-time, happens much less often. A first attempt of interpretation might be that these projects require much more effort in time and technical setting.
- Research that gathers data specifically for the purpose of research is divided in three strategies: scraping (which means to automatically extract information from websites) of large networks or content to examine a status quo, simulating queries against search engines, and surveying users. Scraping (parts of) networks is more frequent than anything else.
- All in all, it is noticeable how low the commonality is to traditional empirical social research, in which user surveys are a mainstay. Obviously, this method is in fact of tremendous importance online. Their minor occurrence here must be due to the fact that research projects in which user inquiries are simply moved to online tools instead of offline tools were filtered out by the rigid definition of web-native (see chapter 1.1). Nevertheless, it can be assumed now that a

significant number of research projects is really web-native in the strictest form of the term.

Apparently, an ontology-based, thorough understanding of all scientific domains related to web-native methods requires considerably deeper investigation of all aspects: One would have to use the ontology for matrices of several interrelations. By doing this, it would for instance be possible to analyse in what way the properties depict domains that were not mentioned in the ontology yet, or evaluate whether the properties are really in the »right place« from a methodological point of view. This would challenge the ontology's intrinsic validity, but could also provide insights into a general methodology of Digital Methods and help to derive a top-down framework of investigating social aspects of the web (»*If you want to research phenomenon y on the web, use web-native method x, which is grounded in the traditional methodology of z, providing means to directly compare new results (of x) and referential studies from the respective domain (z)«).* Alternatively, one could apply a classical (offline) method set of the social research domain and see how much of it is covered here; this would provide more insights in the current state of acceptance of web-native methods as general social research instruments. For now, there is a conceivable correlation of the field of Digital Methods on one hand and the web science's general areas of knowledge about the social, as defined by its founder, Tim Berners-Lee, on the other hand.

More insights into the fields of interests, the methodological set or the development of the scientific domain over time, are provided when analysing the threads (*ResearchDomain* with *SocialResearch*, *Philosophy* and *Politics*, *DigitalMethods*, *TimeFrames*, *ResearchInitiator*) individually, as done in the following subsections.

7.2.1 Digital Methods in the Context of Social Sciences

Previous to a generalization of the web-related social research domain, one might want to have a look at certain segments to predicate statements about detailed aspects. For instance, one would possibly come across the question whether »a past state of the web can be conjured« (see Illustration 7-1) and find its allocation to the domain of

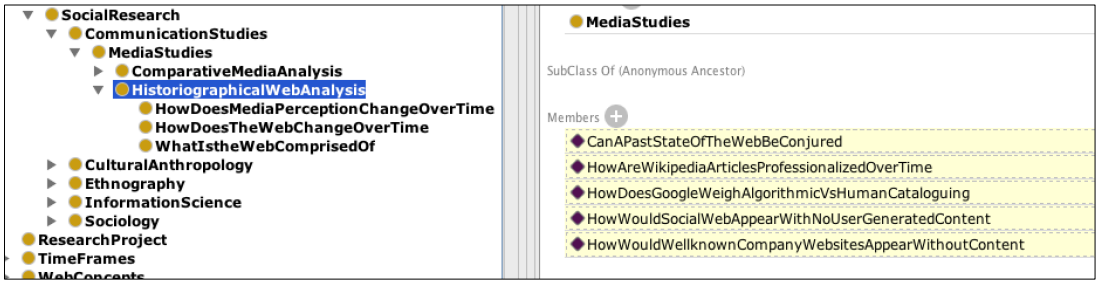

Illustration 7-1: Subclasses of the Domain of Social Research (Protégé screenshot)

MediaStudies inappropriate; instead, the item itself is expectable to be a question of information science. Nonetheless, the class *HistoriographicalWebAnalysis*, in which it is embedded, gives meaning when seeing *all* instances of it: It is a form of web history distinctive from media perception (user focused) and website frames over time stripped of its (user-generated) content. Many of these individual considerations point at the situation of web-related social research as a whole. In this case, it is legitimate to say that the web evolved over time (concerning content, website design, search engines and structure), along the growing interest in user participation and the changes in user behaviour towards the web and particular web services, and that this development can be verified through subordinates of media studies.

In general, social research is represented in the ontology with five major branches:

- Communication studies
- Cultural anthropology
- Ethnography
- Information science
- Sociology

What also shows in the results is that the present ontology did not succeed to include important professional discourses that are present at current state in the discussion about social research in the context of the web. Although Rogers does consider and illustrate why large data sets (»big data«) are difficult for research (Rogers 2013: 201), these considerations are not included in the ontology due to the ontology's limitation on research projects and respective methods. Some concerns have been added during the collection process in a separate class, for instance to illustrate the problem of integrity and privacy of large data sets in the Google Flu Trends research, but other general problems were dismissed due to their lack of conjunction with a certain project. Examples are the problem of how inaccurate web data was in history (»web as space of idiots«) and still is concerning especially social media and user generated content (e.g. due to orthographic mistakes, private opinions), as well as general considerations about hypertext literacy theory and social network theory (ibid. 2013: 27), which should be conciliated with research methods.

7.2.2 Digital Methods in the Context of Philosophy

From the domain of philosophy, only two research projects are mentioned, both concerning the influence on personalization of search engine results as a web-epistemological question.

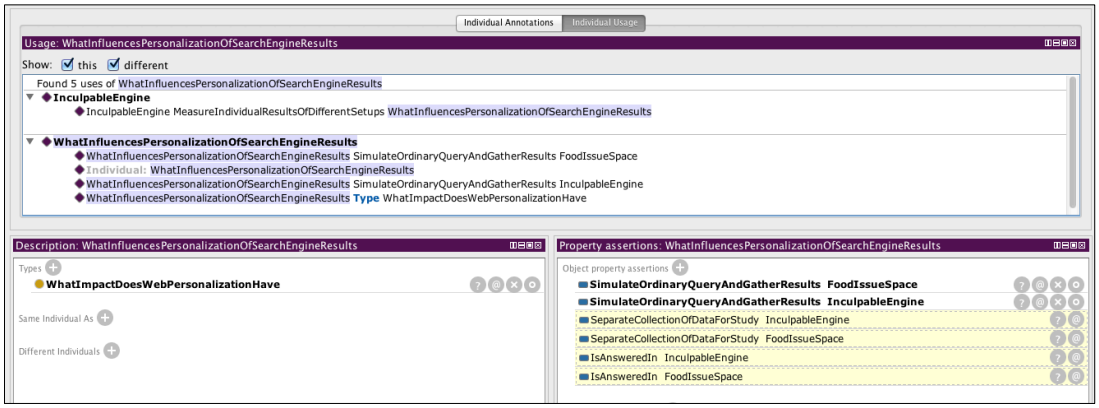

Illustration 7-2: Philosophy Class with two Individuals (Protégé screenshot)

The reason for the absence of more philosophical questions despite the obviously comprehensive research agenda in this field may be grounded in the considerably broad definition of social research of this work (including some philosophical considerations), and the focus that Rogers put on interactive, societal interrogation. Prospective amplification of the Digital Methods ontology with additional domains literally begs for a focus on the *Digital Humanities*, the branch of humanities that is concerned with computerised investigation.

7.2.3 Digital Methods in the Context of Politics

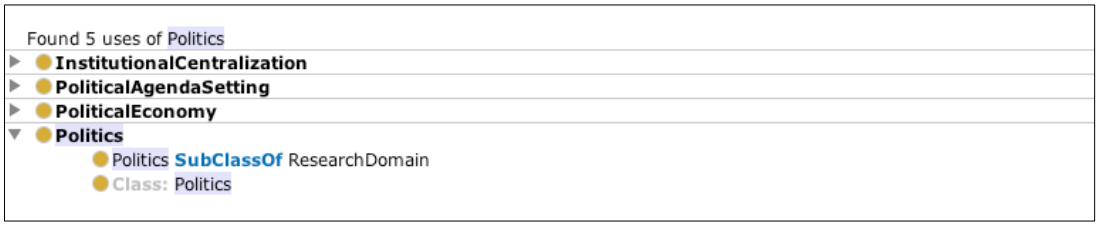

Illustration 7-3: Subclasses of the Politics Class (Protégé screenshot)

Five applications of political research were identified in the present meta study and included as such in the ontology. The area of Institutional Centralization discusses the way institutions associate on the web via links. A comparably ancient project from 1999 attempts to evaluate all outgoing hyperlinks of one source to estimate the influence of offline organizational politics on hyperlink maps (in a quantifiable sense), whereas the second one, conducted only a year afterwards, specifically investigates the *kind* of links that are given and received, and whether they are aspirational, cordial or critical (hence in a qualitative way). All in all, only six individual research projects were associated with the domain of political research (Illustration 7-4). It is important to say that there are more research projects from the field of political science in the ontology,

but they were allocated to the *CulturalAnthropology* class, for reasons illustrated in chapter 6.1. Web censorship research is a very important domain of web-native research, because as opposed to some other phenomena, direct inferences from online to offline situations are valid. Projects described so far concern mainly the Iranian and Iraqi web.

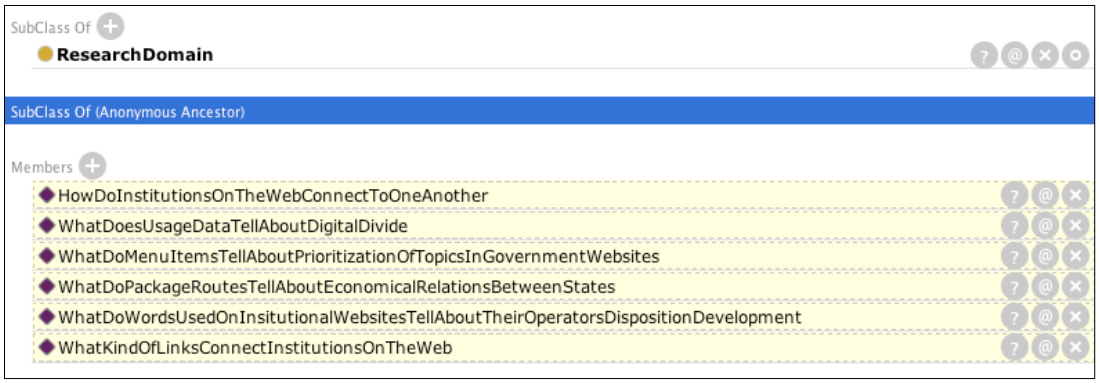

Illustration 7-4: Individuals of the Politics Class (Protégé screenshot)

It becomes apparent that early studies of link politics trying to understand the motivation of linking and construing networks, interpreted linking as strongly correlating with offline networks, e.g. associations of ideology or economical interest or issue driven motivations. Additional motivational paradigms like link impact on search engines, quantity of (social) networks or frequency of releasing communication pieces, came into view later, presumably along with the rising economisation of the World Wide Web.

7.2.4 Digital Methods as an Emerging Empirical Methodology

Within the Digital Methods research domain, six general methodological fields of have been identified, plus a number of methods serving as *Predecessors* for current method proposals (Illustration 5-4).

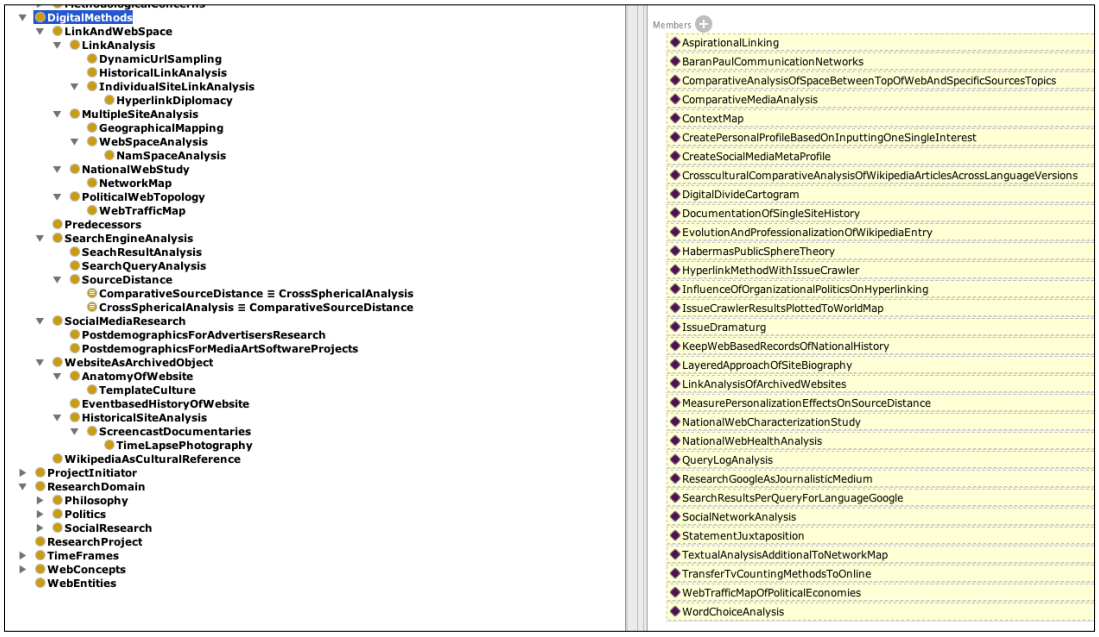

Illustration 7-5: Subclasses and Individuals in the DigitalMethods Superclass (Protégé screenshot)

Nevertheless, not the seven classes offer the most interesting insights, but rather the subproperties of *Utilizes*, which specifies the relation of *ResearchProjects* and respective *DigitalMethods*. Similar to what was done with all subproperties of *Answers*, the forms of utilizing a method to conduct specific studies can be aggregated in cluster:

- Issue tracking (illustrate emphasis on topics over time, discursive floating through networks, language comparison)
- Sampling (estimate world connectivity)
- Social network analysis (gather profile information and spread)
- Network analysis (crawl and detect blocked traffic and routes, web health)
- Word choice analysis (self-censorship, words used over time, word tenor development)
- Comparison of offline and online situations (capture and map states and conditions)
- Search engine analysis (compare result position in trained search engines, over time or with offline occurrences, correlate with offline data and derive conditions)

This simplification indicates that investigating the web with help of web-native methods provides insights concerning public discourse, individual social situations and societal conditions, governmental and structural access limitation, as well as institutional gatekeeping of information.

7.2.5 Digital Methods in the Course of the Years

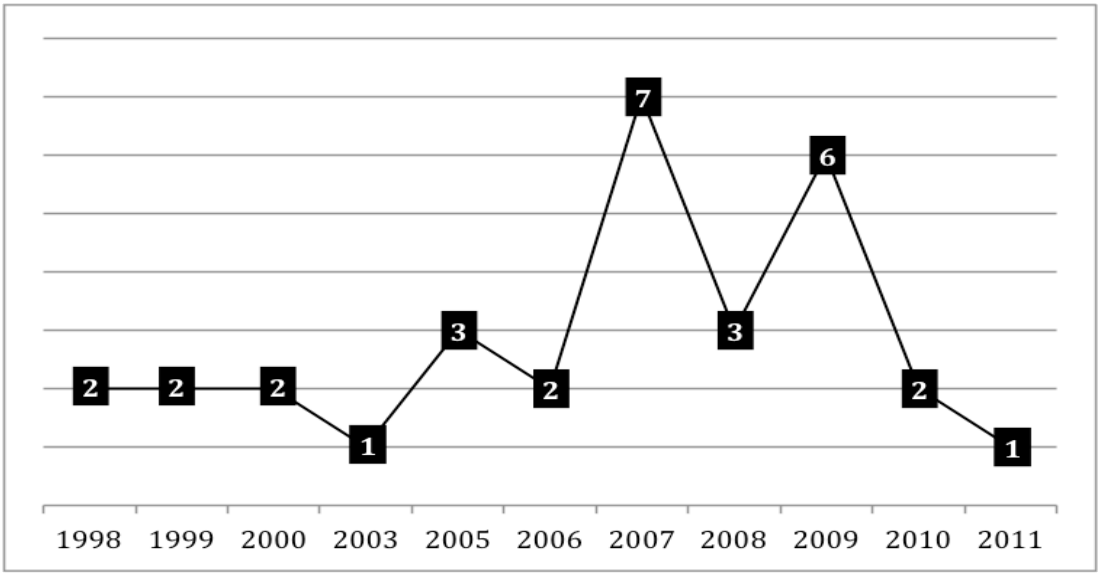

Illustration 7-6: Number of Projects Conducted During Course of Years 1999 - 2011 (own illustration)

Looking at the years of conduction might be interesting to gain insights into trending of issues and transformation of research methods over time, and understand the consecutive development and sophistication of methodology. However, as Illustration 7-6 shows, there is no identifiable increase or decrease. At current state, the ontology is too small for valuable insights of that kind, but prospective additions and expansion might result in more significant numbers.

7.2.6 Digital Methods as a Privilege of the Few?

Similar to how year dates can provide insights into the development of the research field, information about the conductors of all gathered studies might offer valuable knowledge. Two dimensions were examined for each project: What motivation was behind the study, and under which conditions was research conducted? The former led to a distinction of four areas (*EducationalScientificBackground, CommercialBackground, ArtistBackground, JournalisticBack-ground)*, the latter was subdivided into two possibilities (*SinglePerson* or *Institution*), of which the differentiation was quite difficult, as was revealed during the collection process. The ontology shows a clear domination of scientific backgrounds of conductors (15 projects), which is not surprising in this context. It also shows a considerably clear domination of the Digital Methods Initiative (9 projects), which is again not surprising since the unit in question is the initiative of Rogers himself. This number might reveal a topical dominance in the field of Digital Methods, an obvious conclusion about the originator of the term. But, although possible,

this would be an illegitimate assumption drawn, because it might also be an accompanying effect of self-centred collection of examples.

7.3 Conclusion of Interpretation

Overall, some valuable insights into the Digital Methods can be deduced from the ontology. As was shown, the interpretation could solely based on the ontology items provided until now, detached from the initial embedding in the linear book structure. In fact, deriving inferences from the ontology provides a whole new cluster of insights, and allows for totally new questions to be answered due to the possibility to isolate certain areas and analyse them independently from any interferences.

Certainly, at current state, interpretation is limited to a small amount of items that were construed based on only one book, and it is restricted in terms of research efforts. For the future, it would be desirable in a subsequent step to do a deeper analysis of how offline occurrences manifest in the ontology and influence certain spaces. It would for instance be interesting to continue mapping the areas of interests of the Digital Methods (p. 79) to the areas of interest of Berner-Lee et al. (p. 80) and identify differences in their spaces of investigation as well as their methods of researching these spaces. The previously established hypothesis that trust and web morality are areas that have not yet been approached with Digital Methods, would require a second retrieval and classification of relevant research projects. Maybe this would then result in the insight that these research domains can not be researched with web-native data, or that they have already been researched extensively, but were not included in the book. In any case, further investigation would contribute to a better understanding of what happens in the research with web-native data, and what might happen in the future.

For now, it must be postulated that interpretation through generalization is possible, but the validity of statements about the research domain – e.g. about accumulations and gaps of research – would increase significantly along with a future extension of the ontology. In general, any addition and elaboration of the domain, at best in interactive processes with several involved domain experts, would be beneficial for interpretation.

8 Discussion & Conclusion

»Das moderne Denken hat einen beachtlichen Fortschritt gemacht, indem es das Existierende auf die Reihe der Erscheinungen, die es manifestierten, reduzierte« (Sartre 2002).

8.1 Conclusion

The previous chapters have proven that the approach to formalize the knowledge domain of Digital Methods with help of OWL was successful concerning the *correctness of the ontology,* which manifests in the positive evaluation of the three dimensions of quality that have been identified in the beginning of this paper (chapter 1.4) and evaluated in chapter 6:

The intrinsic *result validity* of the ontology is proven, since the Protégé control mechanisms were satisfied and the contentual review in chapter 6.1 removed logical errors; the inductive, bottom-up process was evaluated with help of a control group in chapter 6.3 and revealed no false or invalid approaches to breaking down content into granular units; instead, it proved that the ontology is in fact scalable for future needs. And finally, the requirements concerning user interaction formulated in chapter 2 were satisfied sufficiently, as chapter 6.4 showed.

Apart from this ontology's evaluation that chapter 6 was concerned with, the essential objectives of the present *paper*, derived from the initial research question posed on page 13, open up three more dimensions of success:

1) The desired *improved illustration* of the research field of Digital Methods demanded for a complete *integration* of all concepts known from the book of Richard Rogers.
2) The improved illustration also required *prospective scalability* as a solution to the rapid changes of an emerging scientific discipline, as chapter 1.2 shows.
3) Furthermore, chapter 1.2 introduced the desire for *generalization* as a contribution to the web science domain.

Reflections on these dimensions are provided in the following subsections.

8.1.1 Completeness

Concerning the desired improvement of illustrating the Digital Methods in an ontology, one may think in the concepts of *simple (or efficient)* and *complex (or satisfying).* It can be assumed that the illustration was successful in a *simple* way, because the Digital Methods research field, as described initially by Rogers (2013), could be illustrated *entirely* by means of an ontology. This is proven by the fact that all research projects and methods contained in the book were collected and could be either integrated or disregarded. Both integration and disregard were based on a distinctive definition of relevant and irrelevant items, as illustrated in chapter 3.2, which is why these decisions can be assumed to be unambiguous, uncontroversial and ultimately correct. Additionally to the ontology's formal correctness – ascertained in the last section – this contributes to the perception of a successful transformation of the book format into a formal knowledge representation.

On a *satisfactory* level, a thorough and comprehensive illustration must also prevent coherences and inferences within the knowledge domain, which in the book might be »hidden« in context, from being lost. That is, the dispersion of text into granular information units must not dismiss important information that has only been delivered through context, structure or other experience based performances of readers. That makes it difficult to state that one research project was similar to another if they belonged to different classes. This problem was illustrated and solved in chapter 4.4.

What remains, though, is the challenge of missing prioritisation of ideas over others. It is for instance not possible to state that one research area is generally more important, or more complex and extensive, than another. However, this appears »false« when comparing with the current state of research, knowing e.g. that social media sites currently get much more attention from scientific audiences (of various disciplines) than Wikipedia. The social media thread is almost non-existent right now, though. This problem is known, albeit not perceived as a failure of the ontology for two reasons. Firstly, the taxonomical structure of the resulting ontology can in fact promote one concept over another. It does so by providing different amounts of individuals or deeper ramifications of subclasses. Secondly, this ontology is prepared to grow along the knowledge domain. The more research projects will be conducted about social media, the more complex and important will the respective thread become. This already showed in chapter 6.3, were the random collection of new studies was shown to be almost entirely of a social media subordinate kind.

8.1.2 Scalability

As said before, the ontology awaits adjustments of the domain as a whole and of every contained concept. By adding more research projects, some initial vagueness about the domain, like the missing depth of social media research, will be sharpened. A first step of releasing the ontology to a scientific audience for growth was taken by its integration into the WebProtégé service of the Stanford University, were it can be modified in a web-based environment, providing interaction and mutual agreement on the knowledge domain, as was illustrated in chapter 5.2. Due to the flexible nature of OWL, any other reuse for various purposes is conceivable.

In fact, since OWL is a meta language, its real value lies in the reuse for anything else but a simple being; future scale and use will actually contribute to its improvement in validity, value and self-descriptiveness.

8.1.3 Generalization

Two objectives were formulated in the introduction concerning generalization: On a content-level, a generalizability of results was desired to derive statements about the general field of web-related research from the spotty selection of Rogers' Digital Methods. As chapter 7.2 showed, deductions are already possible at current state, but are not yet based on solid foundation. For instance, so far, this paper refrained from interpreting numbers of studies per year as a quantitative measure for rise or decline in research interests. The more applications of Digital Methods are included in the ontology, the more significant will generalized statements like this become.

A second level concerns the possibility to reuse the process described in this paper – the collection of knowledge pieces and their transformation into an ontology – for other purposes. Although it appears useful to apply the introduced stepwise approach (distinctive definition, collection according to definition, manual cleansing, tripartite evaluation) to similar research intentions, it may not be expedient in any situation. In fact, the resulting process could simply not be *proven* to be universally applicable to other knowledge domains; the major focus of this work was a meta-study of studies about Digital Methods, not to develop a top-down framework of building ontologies. Nevertheless, the nature of ontologies lies in their ability to illustrate *any* knowledge, which is why the process can surely be *repeated* for any other knowledge domain within or outside the field of web science.

8.2 Outlook

In the time between the publishing of the »Digital Methods« book in June 2013 and today, two events took place. With Remote Event Analysis, the Digital Methods Summer School of Amsterdam, hosted among others by Richard Rogers, launched a new branch of Digital Methods in 2014, acknowledging the »growing literature (...) on the relationship between social media and events, often focusing on conflicts, disasters as well as political elections« (Niederer 2014), and analysing what events looked like online and how to systematically follow them. Meanwhile, the ACM Web Science Conference 2013 was held in Paris, and conference proceedings have been published online. Out of 59 published papers, 24(!) were directly or indirectly engaged with web-native methods, asking the very same questions that have been dealt with in this paper: What can be learned from web-native data? Moreover, what can we be learned from the medium »web«? What does web usage reveal about society and culture? How do people behave on the web, and what does that say about offline situations? Although submitted to a web science conference, the academic perspective when researching these social and cultural behaviours is diverse. Among others, the studies have a communicational, an economical and a mobility background. It seems as if the prediction of Lev Manovich came true, who made a plea for cultural analytics back in 2007:

> »We feel that the ground has been set to start thinking of culture as data (including media content and people's creative and social activities around this content) that can be mined and visualized. In other words, if data analysis, data mining, and visualization have been adopted by scientists, businesses, and government agencies as a new way to generate knowledge, let us apply the same approach to understanding culture« (Manovich 2007).

Given that the inseparability of »the offline« and »the online« is proceeding, and a further fusion of research methods for social behaviour online and offline is more than likely in the future, it is more important than ever to provide comprehensible access to its concepts and ideas to as many research professionals as possible. The more interest grows in this branch of web science, the more important are meta-studies that attempt to sort and classify them:

> »Das umfassendere Ziel besteht darin, die Methoden der Internetforschung zu überarbeiten und damit einen neuen Studienzweig zu entwickeln« (Rogers 2011: 62).

Apart from the ease of accessing existing information, this paper may hence itself help in raising the awareness of web-native methods: The more interaction and discussion is evoked about this branch of web science, the better will its character be defined, and the more value will it provide to research.

There are various options for further elaboration of the Digital Methods ontology, and various ideas residing in it to be taken up by other researchers. Some have already been insinuated during this paper. On page 84, a necessity for raising the significance of quantity was identified: Researchers might intend to continue the ontology engineering work by identifying and adding more objects, and successively contribute to possibilities of quantitative, systematic studies. Subsequently, they might for example want to reuse the ontology for perceptions about a chronology of methods, or a cross-sectional study of conductors and motivations of web-native research projects. Concerning the attempt to identify traditional research domains in which the epistemological interest of certain studies may be grounded, a future motivation should be to develop a more reliable and especially more significant approach to schematizing; a proposal for a more generic sorting process would maintain the significance of the ontology in the future. Furthermore, some concepts that are currently objects of tremendous scientific discourse are assumed to be important additions to the ontology in the near future. Besides more work on social network analysis as discussed already, all studies to fall under the catchphrase »Big Data« are to be named here.

Apart from these rather operational objectives, one might already use the current-state ontology for hypotheses about the web in the context of certain domains and research them, since it might already raise some questions that could arouse research interest. For instance, why is the interest of sociology in applying web-native methods rather small? The only documented application of web-native methods by this domain is social network analysis; and even this is seemingly underrepresented within the ontology. This might certainly be due to other (arte)facts, like the focus of Rogers on different research, or an insufficient separation of sociology from other social sciences within the ontology. But further investigation might also reveal that this considerably old, established domain refuses to perceive the web as an »equal« space, and refrains from »digitalizing« its methodology. The ontology itself will provide no answer to this, but can serve as a starting point for deeper investigation. One might also take a closer look at how the only question asked by web epistemology concerns the impact that personalization effects have on search engine results – instead of going beyond to ask how the hyper-personalisation of information through online technologies shapes the way individuals perceive the world. Again, the ontology may serve as an instigator for further research.

9 Resources

Altmeppen, K., Weigel, J. & Gebhard, F. (2011). „Forschungslandschaft Kommunikations- und Medienwissenschaft." *Publizistik – Vierteljahreshefte für Kommunikationsforschung*, November 8, 2011: 374-398.

An, J., Quercia, D., Cha, M. et al. (2013). Traditional media seen from social media. Proceedings of the 5th Annual ACM Web Science Conference on - WebSci '13, 11–14. doi:10.1145/2464464.2464492

Anticoli, L. & Toppano, E. (2013). Technological Mediation of Ontologies: the Need for Tools to Help Designers in Materializing Ethics, 1(3), 23–31. Retrieved from http://www.seipub.org/ijps/PaperInfo.aspx?ID=11159

Benninghaus, H. (1998). Deskriptive Statistik. 8th Edition. Bd. 1. 4 Bde. Stuttgart; Leipzig: Teubner, 1998.

Berners-Lee, T., Hall, W., Hendler, J. et al. (2006a). „A Framework for Webscience." *Foundations and Trends in Web Science*, Nr. Vol. 1 (2006): 1 - 130.

Berners-Lee, T., Hall, W., Hendler, J. et al. (2006b). „Creating a Science of the Web." *Webscience Trust.* August 11, 2006. http://journal.webscience.org/2/2/creating.pdf (accessed November 1, 2013).

Borra, E. (2014). About the Digital Methods Initiative. March 20, 2014. https://wiki.digitalmethods.net/Dmi/DmiAbout (accessed April 8, 2014)

Breitman, K., Casanova, M. & Truszkowski, W. (2007). *Semantic Web - Concepts, Technologies and Applications.* London: Springer, 2007.

Brosius, H., Haas, A. & Koschel, F. (2012). „Methoden der empirischen Kommunikationsforschung." *Studienbücher zur Kommunikations- und Medienwissenschaft.* Wiesbaden: Springer Fachmedien, 2012.

Bruns, A. (2014). „#Ausvotes: Twitter Activity Across the Electorates." *Mapping Online Publics.* August 22, 2013. http://mappingonlinepublics.net/2013/08/22/ausvotes-twitter-activity-across-the-electorates/ (accessed March 18, 2014).

Bush, V. (1997). *As We May Think.* Bd. 2, in *FormDiskurs – wiedergelesen/Re-read*s, von Hartmut Winkler, 136-147. Frankfurt a.M.: form, 1997.

De Choudhury, M., Counts, S., & Horvitz, E. (2013). Social media as a measurement tool of depression in populations. *Proceedings of the 5th Annual ACM Web Science Conference on - WebSci '13*, 47–56. doi:10.1145/2464464.2464480

Debin, M., Souty, C., Turbelin, C. et al. (2013). Determination of French influenza outbreaks periods between 1985 and 2011 through a web-based Delphi method. *BMC Medical Informatics and Decision Making*, *13(1)*, 138. doi:10.1186/1472-6947-13-138

Dede, C. (2008). „A Seismic Shift in Epistemology." *Educause Review Online.* June 2008. http://www.educause.edu/ero/article/seismic-shift-epistemology (accessed March 6, 2014).

Fernback, J. (1999). „There is a there there: Notes toward a definition of cybercommunity." In *Doing internet research: Critical issues and methods for examining the net*, von Steve Jones. Thousand Oaks, California: SAGE Publications, 1999.

Gloria, M. Difranzo, D., Fernando, M. et al. (2013). The Performativity of Data : Re-conceptualizing the Web of Data, 109–117.

GoodRelations Wiki (2013). *Prominent Users of GoodRelations.* April 9, 2013. http://wiki.goodrelations-vocabulary.org/References (accessed January 18, 2014).

Horridge, M. (2011). *A Practical Guide To Building OWL Ontologies Using Protege 4 and CO-ODE Tools* . Prod. University of Manchester. Manchester, March 2011. Retrieved from http://home.skku.edu/~samoh/class/sw/ProtegeOWLTutorial.pdf

Hunt, E. & Colander, D. (2014). *Social Science: An Introduction to the Study of Society.* 14th Edition. New Jersey: Pearson, 2014.

Jonassen, D., Collins. M, Davidson, M. et al. (1995). „Constructivism and computer-mediated communication in distance education." *The American Journal of Distance Education*, 1995: 7-26.

Krug, S. (2000). Don't make me think. A Common Sense Approach to web usability. San Francisco: New Riders Publishing.

Legrady, G. (2014). *Making Visible the Invisible (2005 - 2014).* March 29, 2014. http://www.mat.ucsb.edu/g.legrady/glWeb/Projects/spl/spl.html (accessed March 29, 2014).

Lima, M. (2014a). *Top 10 Twitter Languages in London.* March 29, 2014. http://www.visualcomplexity.com/vc/project_details.cfm?id=777&index=777&domain= (accessed March 29, 2014).

Lima, M. (2014b). *Wikipedia Articles During the Middle-East Protests.* March 29, 2014. http://www.visualcomplexity.com/vc/project_details.cfm?id=765&index=765&domain= (accessed March 29, 2014).

Manovich, L. (2007). Cultural Analytics: Analysis and Visualization of Large Cultural Data Sets. In *Software Studies Initiative*. Retrieved from http://manovich.net/cultural_analytics.pdf

Mapping Online Publics (2014). „About." *Mapping Online Publics.* March 18, 2014. http://mappingonlinepublics.net/about/ (accessed March 18, 2014).

McGuinness, D. & van Harmelen, F. (2004): OWL Web Ontology Languge Overview (W3C Recommendation). http://www.w3.org/TR/owl-features/ February 10, 2004. http://www.w3.org/TR/owl-features/ (accessed April 11, 2014)

Niederer, S. (2014). *Call for Participation - Digital Methods Summer School 2014.* 27. March 2014. https://wiki.digitalmethods.net/Dmi/SummerSchool2014 (accessed March 31, 2014).

Noy, N., & McGuinness, D. (2000). Ontology Development 101 : A Guide to Creating Your First Ontology, 1–25. Retrieved from http://ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness.pdf

Oxford Internet Institute (2012). *The Geographically Uneven Coverage of Wikipedia.* November 2012. http://geography.oii.ox.ac.uk/?page=the-geographically-uneven-coverage-of-wikipedia (accessed March 29, 2014).

Parsons, T. (1967). An Approach to the Sociology of Knowledge. In *Sociological Theory and Modern Society* (pp. 139–165). New York: Free Press. Retrieved from http://solomon.soth.alexanderstreet.com/cgi-bin/asp/philo/soth/getdoc.pl?S10019968-D000006

Parsons, T. & Shils, E. (2001). *Toward a General Theory of Action – Theoretical Foundations for a Social Science.* Originally published in 1951. New Brunswick, New Jersey: 2001.

Pickering, R. (2014). *The Music Ontology.* 18. January 2014. http://musicontology.com/ (accessed January 18, 2014).

Protégé Wiki (2014). *OWLViz.* 23. July 2013. http://protegewiki.stanford.edu/wiki/OWLViz (accessed January 14, 2014).

Rogers, R. (2011). „Das Ende des Virtuellen." *Zeitschrift für Medienwissenschaft*, 2011: 61-77.

Rogers, R. (2013). *Digital Methods.* Cambridge: The MIT Press, 2013.

Sartre, J. (2002). *Das Sein und das Nichts – Versuch einer phänomenologischen Ontologie.* 8. Edition. Reinbek bei Hamburg: Rowohlt Taschenbuch Verlag, 2002.

Scherfer, K. & Volpers, H. (2013). „Einführung." In *Methoden der Webwissenschaft*, by Scherfer, K. & Volpers, K. (ed.), 7-12. Münster: LIT-Verlag, 2013.

Stanford University (2014a). „Protégé Introduction." *Protégé Resources.* March 2014. http://protege.stanford.edu/ (accessed March 27, 2014).

Stanford University (2014b). „WebProtégé Introduction." *Protégé Resources.* March 2014. (accessed March 27, 2014).

van Dijk, T. (1980). *Textwissenschaft: eine interdisziplinäre Einführung.* German translation: Christoph Sauer. Tübingen: Niemeyer, 1980.

W3C (2013). *Good Ontologies.* December 13, 2013. http://www.w3.org/wiki/Good_Ontologies (accessed January 18, 2014).

W3C Working Group (2009). „OWL Web Ontology Language – Use Cases and Requirements." *World Wide Web Consortium w3.org.* November 12, 2009. http://www.w3.org/TR/webont-req/ (accessed January 13, 2014).

W3C Working Group (2012): OWL 2 Web Ontology Language Overview (W3C Recommendation). December 11, 2012. http://www.w3.org/TR/owl2-overview/ (accessed April 11, 2014)

Wikipedia (2013a). *Cultural Studies.* November 19, 2013. http://en.wikipedia.org/wiki/Cultural_studies (accessed January 27, 2014).

Wikipedia (2013b). *Kommunikationswissenschaft.* December 26, 2013. http://de.wikipedia.org/wiki/Kommunikationswissenschaft (accessed January 27, 2014).

Wikipedia (2013c). *Sozialwissenschaften.* Oktober 8, 2013. http://de.wikipedia.org/wiki/Sozialwissenschaften (accessed January 27, 2014).

Wikipedia (2014a). *Epistemology.* January 23, 2014. http://en.wikipedia.org/wiki/Epistemology (accessed January 24, 2014).

Wikipedia (2014b). *Ethnographic.* January 18, 2014. http://en.wikipedia.org/wiki/Ethnographic (accessed January 27, 2014).

Wikipedia (2014c). *Ethnomethodology.* March 4, 2014. http://en.wikipedia.org/wiki/Ethnomethodology (accessed March 16, 2014).

Wikipedia (2014d). *Information Science.* February 28, 2014. http://en.wikipedia.org/wiki/Information_science (accessed March 16, 2014).

Wikipedia (2014e). *Media Studies.* January 15, 2014. http://en.wikipedia.org/wiki/Media_studies (accessed January 27, 2014).

Wikipedia (2014f). *Medienwissenschaft.* January 11, 2014. http://de.wikipedia.org/wiki/Medienwissenschaft (accessed January 11, 2014).

Wikipedia (2014g). *Political Science.* January 16, 2014. http://en.wikipedia.org/wiki/Political_science (accessed January 27, 2014).

Wikipedia (2014h). *Politics.* March 11, 2014. http://en.wikipedia.org/wiki/Politics (accessed March 16, 2014).

Wikipedia (2014i). *Social Network.* March 14, 2014. http://en.wikipedia.org/wiki/Social_network (accessed March 16, 2014).

Wikipedia (2014j). *Social Science.* March 12, 2014. http://en.wikipedia.org/wiki/Social_science (accessed 16. March 2014).

Wikipedia (2014k). *Sociology.* January 12, 2014. http://en.wikipedia.org/wiki/Sociology (accessed January 27, 2014).

Wikipedia (2014l). *Workplace Politics.* January 21, 2014. http://en.wikipedia.org/wiki/Workplace_politics (accessed March 16, 2013).

Wikipedia (2014m). *Anthropologie*. March 30, 2014. http://de.wikipedia.org/wiki/Anthropologie#Geisteswissenschaftlicher_Ansatz (accessed April 3, 2014).

Declaration in Lieu of Oath

I hereby declare that this master thesis was independently composed and authored by myself.

All content and ideas drawn directly or indirectly from external sources are indicated as such. All sources and materials that have been used are referred to in this thesis. The thesis has not been submitted to any other examining body and has not been published.

Cologne, April 2014

Miriam Schmitz