Conceptual Modeling as Language Design

Ryan Shaw

School of Information and Library Science

University of North Carolina at Chapel Hill

Author Note

Ryan Shaw iD https://orcid.org/0000-0002-5486-159X

I have no conflicts of interest to disclose.

Correspondence concerning this article should be addressed to Ryan Shaw, School of Information and Library Science, CB #3360, 100 Manning Hall, Chapel Hill NC 27599-3360, USA. Email: ryanshaw@unc.edu

Abstract

Conceptual modeling is a broad practice encompassing knowledge organization, domain modeling, and knowledge representation. It is best understood not as a scientific process of discovery, but as a constructive process of language design. This constructive process involves both explicating differences of meaning implicit in some discursive tradition and revising those differences to improve that tradition. Understood this way, conceptual modeling can serve as the basis for a paradigm of information research and practice that does not reproduce fundamental asymmetries between researchers and the people they study, or between practitioners and the people they serve. Progress within this paradigm will involve combining methods from what have up to now been different traditions or modes of conceptual modeling. It will require paying closer attention to the historical and structural dimensions of discursive traditions and reimagining the function of critique.

Conceptual Modeling as Language Design

Conceptual models make the world intelligible. In that sense, any theory is a conceptual model. But only some theories are materialized as formal structures, and it is such formal structures that I focus on here. Formally structured conceptual models can be found everywhere in information systems. Knowledge organization systems such as thesauri or classification schemes are conceptual models, the formal structures of which function as abstract spaces for locating items of recorded information. Software designers and engineers create domain models, the formal structures of which organize both the software that implements information systems and the labor of creating them. So-called artificial intelligences are comprised of knowledge representations, the formal structures of which define what *intelligence* is supposed to mean in some context. All these kinds of conceptual models are theories that create certain possibilities and rule out others. Whether one can refute these theories is irrelevant, at least to the extent that one relies on such information systems to live.

For a long time, these kinds of formally structured conceptual models were the remit of various specialist professions. Discussions centered largely on methods for constructing conceptual models and how to evaluate them. No consensus on those methods has been reached since these specialist professions are mostly disconnected from one another. Occasionally these specialized discussions turned to reflections on the consequences of conceptual modeling for broader society, but these were few and far between. Recently, however, as large-scale, networked information systems have begun to play an increasingly important role in contemporary life, more people have begun to pay attention to the models or "algorithms" that are realized by these systems. Often, that attention takes the form of critique; less often, that critique is informed by a thorough understanding of what conceptual modeling is. But if some critics lack understanding of conceptual modeling, the difficulty that conceptual modeling specialists have had responding to these critiques have demonstrated that their own understanding of what they do is far from complete.

In this paper I make a case for understanding conceptual modeling as language design. I begin by identifying three partially overlapping but largely disconnected traditions of conceptual modeling. I show how each of these traditions includes some who understand conceptual modeling to be capturing an impartial picture of an objective reality, and others who understand it to be designing an arbitrary language to support situated action. Rather than arguing that the latter view is correct, I instead develop it in detail, drawing upon the work of the scholar of librarianship Vesa Suominen and the philosopher Rudolf Carnap. Building upon Suominen's rationale for understanding language design as a paradigm for librarianship, I argue that conceptual modeling can serve as the basis for a paradigm of research and practice in the information field more broadly. I end with a discussion of why I believe this paradigm would be a fruitful one, and some of the specific research challenges it implies.

Different Traditions of Conceptual Modeling

There is no unified tradition of conceptual modeling. Instead, there is a complex genealogy of partially overlapping traditions, each with its own preferred terminology, tools, and techniques. One such tradition treats conceptual models as *maps for navigating a space of topics*. Professionals working in this tradition aim to develop tools that can inform about concepts of interest in some domain, various names or terms associated with those concepts, and relationships among concepts (Hjørland, 2007). Examples include taxonomies, thesauri, gazetteers, and subject classifications. These kinds of knowledge organization systems are conceptual models, intended to be used by professionals when (usually manually) cataloging, indexing, and classifying documents. Because such models specify positions in an abstract space to which documents can be assigned (Wilson, 1968, Chapter IV), they are useful not only for finding documents but also for getting a sense of the overall space within which they have been located. This is the tradition of conceptual modeling with roots in the practice and theory of librarianship. I will refer to this tradition as *conceptual modeling in the mode of knowledge organization (KO)*.

A second tradition treats conceptual models as *blueprints for information system construction*. Professionals working in this tradition aim to develop formal models to guide the planning, implementation, and ongoing functioning of information systems. This practice goes by many different names: data modeling, business modeling, systems analysis, document engineering, etc. Conceptual modeling in this mode typically involves the use of tools such as the Unified Modeling Language, entity-relationship diagrams, or data dictionaries. The models created are intended to simplify and structure knowledge about the domain to which an information system will be applied, to serve as the basis for communication between system developers and domain experts, and to ensure that the implemented system can be interpreted in terms of domain concepts (Evans, 2004, pp. 3–4). This is the tradition of conceptual modeling typically taught to professional students in information science, management of information systems, and software engineering. I will refer to this tradition as *conceptual modeling in the mode of domain modeling (DM)*.

A third tradition treats conceptual models as *mental representations for thinking machinery*. Professionals working in this tradition aim to develop formal representations of the world that can support automated reasoning. This tradition also goes by various names: knowledge representation, knowledge engineering, formal ontology, machine learning, etc. This tradition is rooted in the research program of artificial intelligence (AI). One important branch was pioneered by Edward Feigenbaum, who in the 1970s shifted the research focus in AI away from algorithms for reasoning toward the representation of knowledge in knowledge bases and expert systems. However, not only so-called symbolic AI but also the currently predominant forms of statistical AI can be understood as the engineering of knowledge representations. A knowledge

representation is a conceptual model that serves some of the same purposes as a domain model, but also functions as "a fragmentary theory of intelligent reasoning" (Davis et al., 1993, p. 17). It represents not only the domain but also a recommended process for (automating) thinking about it. I will refer to this tradition as *conceptual modeling in the mode of knowledge representation (KR)*.

Two Views on Conceptual Modeling

Despite somewhat different conceptions of their goals, each of these traditions draws upon a base of existing documentation and attempts to make it tractable through a process of abstraction and formalization, or *modeling*. But what exactly does conceptual modeling attempt to model? How should this practice be understood? One view of conceptual modeling is that it is essentially a descriptive practice. According to this view conceptual modeling is a kind of "reality mapping" (Lyytinen, 1987, pp. 10–14). The goal of reality mapping is to produce an impartial picture of an objective reality. Conceptual models are factual descriptions of that reality. They represent what everyone must agree on: a shared worldview. Looking at conceptual modeling in the mode of KO, Furner (2012) identified this view with a realist orientation to aboutness. For realists, subjects exist independently of what people think and do. To say that a work is about some subject is simply to recognize the relation that exists between the work and the independently existing subject. Thus, informed observers should be able to reach a consensus on the subjects that a given work is about. Furthermore, based on the relations of aboutness between subjects and works it should be possible to establish a corresponding classification of works by subject. Such a classification would be a map of the actually existing subjects that the works in some domain are about.

Simsion (2007) found a similar perspective in his study of professionals who do conceptual modeling in the mode of DM. Some professionals described their work as largely descriptive. They viewed domain modeling as a process of discovering actually existing concepts implicit in work practices and documentation. For these professionals, domain models were representations of reality—selective and simplified, but nonetheless discovered and not designed. The reality-mapping perspective can also be found among conceptual modelers in the mode of KR. In the realm of symbolic KR, some researchers (e.g., Smith, 2004) have argued that knowledge representations should reflect scientific consensus on the specific things and kinds of things that really exist. From these reality mappers' perspective, good knowledge representations are context-independent and should only change if scientific consensus turns out to have been incorrect. A similar perspective can be found among some proponents of statistical AI, who have defined their goal as the creation of intelligent systems that can learn the "underlying causal properties of the world" from audio, images, and video (Bengio et al., 2021, p. 64). The implication is that the knowledge representations learned by such systems should map reality itself, not just reality as represented in their audiovisual input.

A different view of conceptual modeling is that it is essentially a constructive practice. According to this view conceptual modeling is a kind of language design (Lyytinen, 1987, 14–17). The goal of language design is to support action within some context. Understood as a language, a conceptual model captures (some of) the social conventions and institutions that provide order in some context. It is not a fundamental agreement about the nature of reality, but a temporary compromise, always subject to renegotiation. Those KO researchers and practitioners who Furner (2012) identifies as aboutness nominalists see subject classification as a kind of language design. Subjects are "merely linguistic expressions that serve as labels or names for sets of works" (p. 511). These expressions are conventions that organize communication about works. Using these expressions does not commit one to a precise shared definition of the subject; on the contrary the practical value of such expressions is that they enable communication without having to make such commitments (Stalnaker, 1967). Thus, from a nominalist perspective a conceptual model is not a map of reality but a language enabling communication about works within a context. It is a language that should be expected to change as that context does.

Simsion (2007) also found domain modelers who pushed back against the idea that their work is merely descriptive; these professionals viewed themselves as designers. They saw themselves as actively forging effective compromises among people with different perspectives and goals, not simply writing down "what everyone knows." For these designers, conceptual modeling involved finding the right means of expression to support communication, not discerning the outlines of reality. Some KR researchers (e.g., Gruber, 1995; Dumontier & Hoehndorf, 2010) have similarly emphasized that knowledge representations are languages intended to support communication, including communication with or through "intelligent" machines. Like other languages, they have argued, knowledge representations must be able to accommodate different perspectives. This view is less in evidence among researchers who embrace the research program of AI, perhaps because they view themselves as designing minds, not languages. Still, some forward-thinking researchers have argued for a vision of AI as the design of languages for communicating with alien minds (Donahue, 2021).

Suominen's Semiology of Conceptual Models

Suominen (1997, 2016) provides a compelling account of conceptual modeling as a metasemiotic process of language design. His account focuses on KO (which he refers to as *documentation*), but it is general enough to describe conceptual modeling in all the three traditions identified above. Suominen starts with the phenomenon of *documentary communication*, ongoing discursive practices mediated by documents. A document is any "message with some permanence" (Suominen, 1997, p. 57); others might say *recorded information*. The law is mediated by briefs, court transcripts, and decisions; science is mediated by lab notebooks, journal articles, monographs, CVs, and peer reviews; art is mediated by gallery brochures, auction catalogs, museum collections, and coffee table books; business is mediated by

contracts, invoices, memos, and press releases; and so on. Agre (1995) dubbed such configurations of document use *institutional circuitries*, "the forms and pathways that specific social formations maintain for the movement of their own categories of communicative practice" (p. 227). Both Agre and Suominen distinguish between 1) the subject matter of some discursive practice, 2) the history of the social formation or institution sustained through that practice, and 3) the forms and pathways of documentary communication maintained by that institution.

To participate effectively in some process of documentary communication, individuals must develop *communicative competence*. Suominen borrows the notion of communicative competence from sociolinguistics, where the term is generally attributed to the linguist and anthropologist Dell Hymes (1972; Cazden, 2011). An individual's communicative competence is their ability to draw upon a shared reservoir of features, strategies, and techniques employed in some discursive practice. Different individuals have different degrees of communicative competence, and one's competence may increase over time; increasing communicative competence is a major goal of education and part of acculturation into any field or pastime. Hymes (1972) identifies three aspects of communicative competence. The first aspect is one's *communicative repertoire*; this includes things like the words one knows but also things like one's ability to draw diagrams or to shoot and edit video. The second aspect is knowledge of *communicative routines*, which identify common patterns of communication such as *job interview* or *cocktail party banter*. Communicative routines include the forms of documentary communication focused on by Agre and Suominen; Hymes gives the example of literary genres. The third aspect of communicative competence is understanding which forms of communication are appropriate in particular contexts or situations, what Hymes calls *domains of communicative behavior*.

Suominen's notion of competence with respect to documentary communication includes all three of these aspects. To engage effectively in documentary communication, one must have some knowledge of what the documents are about, i.e. the portions of reality outside the documents to which they refer; the extent of such knowledge determines the limits of one's ability to "speak the language" of some domain. Understanding communicative routines and domains of communicative behavior requires familiarity with the history of the social institutions that are sustained through those routines and establish those domains. So, for example a competent scholar should know how to talk about not only of the subject matter of their field, but also its conceptual genealogy: what the different scholarly traditions are that people see themselves as working within or against. Ideally, communicative competence also includes knowledge of the specifically documentary aspects of institutional history. Again, using scholarship as an example, this would include changes in publication practices, shifting conventions for citation, the background and evolution of key journals or conferences, etc. It is expertise in such applied

history—especially documentary history—that Suominen sees as characteristic of his ideal of librarianship.

Librarians, bibliographers, archivists, and kindred professionals can employ their communicative competence to help others develop their own communicative competence. An individual's communicative competence is the degree to which they have mastered a discursive practice, where mastery consists of an intuitive understanding of a system of differences manifested at all the levels of discourse discussed above. But Suominen (1997) insists that communicative competence should be "conceived of not merely as the mastering of some systems of expressions, but more deeply, as having an appropriate system of (differences of) content in use" (p. 82). A language, as a shared reservoir of communicative possibility potentially available to individuals, is a system of differences of content in use. Participants in a communicative practice need not content themselves with the language they have inherited: they can design new ones. Suominen calls such designed languages *documentary languages*. A documentary language is an artificial language intended to provide professionals with "a system of differences of content needed … to create communicational competence for 'talking about documents'" (Suominen, 1997, p. 138). It is a kind of "practical theory" about the structure of the history of documentary communication in some domain (Suominen, 1997, p. 183).

Conceptual modeling is, in Suominen's terms, documentary language design (so henceforth I will refer to documentary languages as conceptual models). Classification schemes, subject heading lists, and thesauri are examples that Suominen gives of conceptual models, but as explained above the class of conceptual models is not limited to these familiar KO tools. Various kinds of domain models and knowledge representations can also be understood as conceptual models under Suominen's account. What makes these different kinds of things conceptual models, despite the widely varying forms they take, is that each provides specialized means of communicating about communication via documents (much as linguistics provides us with specialized means of communicating about communication via language). This specialized means of communication is structured as a formal system of differences. According to structuralist semiotic theories, such a system of differences does not simply give names to already existing concepts; rather both the names (expressions) and concepts (content) are *created* by the structure manifest in the system of differences (Suominen, 1997, pp. 45–46). This differentiation of meaning involves two steps: partitioning into units of a syntagmatic process, followed by articulation of those units into a paradigmatic system.

The semiotic terminology may be unfamiliar to some, but the process of differentiation described should be familiar to anyone with experience designing conceptual models. Consider for example the design of a tool for faceted browsing of healthcare records. Partitioning the substance of communication about healthcare records into syntagmatic units involves identifying the various aspects of those records that someone might be interested in, such as: the age and

ethnicity of the patient, their medical history, recent lab test results, and what drugs they are taking. Each of those units (which will become a facet or group of related facets in the tool) may be further partitioned; for example, lab test results might be further partitioned into when the sample was collected, when the test was processed, quantitative results, and qualitative assessment. Once this partitioning has been done, the resulting syntagmatic units can then be articulated into a paradigmatic system of exhaustive and mutually exclusive values. Qualitative test assessments might be articulated into *high*, *medium*, and *low*; test processing times might be articulated into ISO 8601 dates; patient ethnicity might be articulated into *American Indian or Alaska Native*, *Asian*, *Black or African American*, etc. Underlying the faceted browsing tool is a conceptual model that divides up reality in some definite but arbitrary way; it is a product of collective human choices that might have been made differently.

Though Suominen does not discuss it, statistical machine learning can also be understood as the formal differentiation of meanings implicit in a discursive practice. For example, state of the art methods for word sense induction (e.g., Amrami & Goldberg, 2018) will begin with the automatic identification of common patterns of syntagmatic units in large text corpora (Hearst, 1992). These patterns then serve as the basis for articulating systems of paradigms (word senses) using probabilistic clustering. The structure of the resulting conceptual model is very different from that of the faceted browsing tool—with "soft" statistical distributions replacing sharply delineated classes—but the process of differentiation is the same. Nor is the differentiation process limited to textual discourse; generative adversarial neural networks have proven to be a flexible technique for finding syntagmatic structural patterns in collections of images and audio as well as text (Goodfellow et al., 2014). When a neural network trained to generate images is also trained on textual descriptions of images, the syntagmatic visual patterns can be articulated into precise models of visual paradigms such as *Unreal Engine*, *ultra high definition desktop wallpaper*, or *Tim Burton* (Shane, 2021).

Carnap's Ideal of Explication

Suominen uses the term *explication* to refer to the process of differentiation that leads to a conceptual model. An explication attempts to formally capture the distinctions of meaning needed in some specific context or for some specific purpose. By formally capturing and making explicit distinctions implicit in some shared reservoir of communicative potential, an explication serves as a tool for improving communication and reaching a common understanding. Suominen (indirectly) takes the concept of explication from the philosopher Rudolf Carnap.[1] Carnap is sometimes presented as an extreme scientific reductionist who dreamt of reconstructing all human knowledge on a sound foundation of empirical observations expressible using formal logic. It is possible to interpret some of Carnap's early work in that light. But Carnap's ideas evolved over time, and in recent years philosophers have re-evaluated his later work to present a more nuanced picture of his thought (Carus, 2007; Leitgeb & Carus 2021).

Much of this re-evaluation has focused on Carnap's notion of explication. Carnap (1950) defined explication as "the transformation of an inexact, prescientific concept … into a new exact concept" (§2). A successful explication should result in a new concept that is like the one it succeeds but is more precise, and that is useful for stating generalizations but is as simple as possible (Carnap, 1950, §3). Contrary to his reputation as a reductionist, Carnap (1950) did not believe that "correct" explications could be discovered scientifically:

> … if a solution for a problem of explication is proposed, we cannot decide in an exact way whether it is right or wrong. Strictly speaking, the question whether the solution is right or wrong makes no good sense because there is no clear-cut answer. The question should rather be whether the solution is satisfactory, whether it is more satisfactory that another one, and the like. (§2)

In other words, explication is not a matter of science but of engineering—or design. Carnap's notion of explication as the design of better conceptual tools is right in line with Suominen's account of documentary languages as tools for coming to grips with the history of documentary communication in some area.

Suominen's account of conceptual modeling emphasizes the development and maintenance of individual communicative competence. That is to be expected, given his rationale for librarianship as a profession focused on educating people about the relevant distinctions that ought to be understood when engaging with the documentary history of some field. This emphasis is also appropriate for conceptual modeling in the mode of KO. Someone with a high degree of communicative competence with respect to some body of literature may benefit less from KO tools than someone new to the field, unless it is a particularly fast-changing one. Carnap, on the other hand, emphasizes the potential for conceptual modeling to improve communicative capacity at a *collective* level. This is a more ambitious program for conceptual modeling, one not content to simply explicate distinctions already in use, but also introducing *new* distinctions enabling better thinking and communication. Carnap's emphasis on concept revision is perhaps more appropriate for conceptual modeling in the mode of DM and, especially, KR, both of which (in comparison to KO) more commonly focus on making new forms of communication possible.

But this is just a difference in emphasis; both Carnap and Suominen see explication as an essentially normative process of choosing which essentially arbitrary distinctions are best suited to some particular purpose, and that there should be different conceptual models in any given toolkit. Suominen stresses that any system of differences resulting from a process of explication might have been been structured some other way. Therefore, a conceptual model must be understood as an independent cultural form, the product of collective human choices, something to be interpreted and understood in terms of history and structure (Suominen, 1997, p. 48).

Suominen emphasizes the constructive nature of explication by identifying it as an example of what Barthes (1972) calls "the structuralist activity":

> … a veritable fabrication of a world which resembles the primary one, not in order to copy it but to render it intelligible. Hence one might say that structuralism is essentially *an activity of imitation*, which is also why there is, strictly speaking, no technical difference between structuralism as an intellectual activity, on the one hand, and literature in particular, art in general, on the other … (p. 215)

The aim of explication is not to faithfully represent the world, but to decompose and recompose it to render it intelligible.

Conceptual Modeling as the Basis for an Alternative Paradigm of IS/S

Within the field of information science/studies (IS/S), conceptual modeling is typically treated as part of the subfields of KO or systems analysis, of little interest to researchers and practitioners working outside of those subfields. But Suominen and Carnap's accounts of conceptual modeling are ambitious enough to qualify as an alternative paradigm for the field as a whole. The standard paradigm of information science (Figure 1) posits that there are various ongoing communication processes involving people, technology, and information. Information scientists are tasked with observing these processes, paying special attention to the form and organization of information and how these relate to the goals of the people creating and using it. Because participants in the processes are preoccupied with the content of information, they lose the ability to explicitly describe its form and organization. So, information scientists should stand outside the processes they observe, and from that standpoint they will be able to discern and explicitly describe the form and organization of information without being distracted by its content (Bates, 1999). From this meta-level, information scientists produce generalized conceptual models. When working in a pure scientific mode, information scientists use their models to explain and predict information behavior. When working in an applied mode, they use their models to inform the design of information technologies that optimize the efficient use of information (Borko, 1968). According to the standard paradigm, information science maps a substrate of form and organization of information invisible to those who merely *know* information.

Figure 1

The Standard Paradigm of Information Science
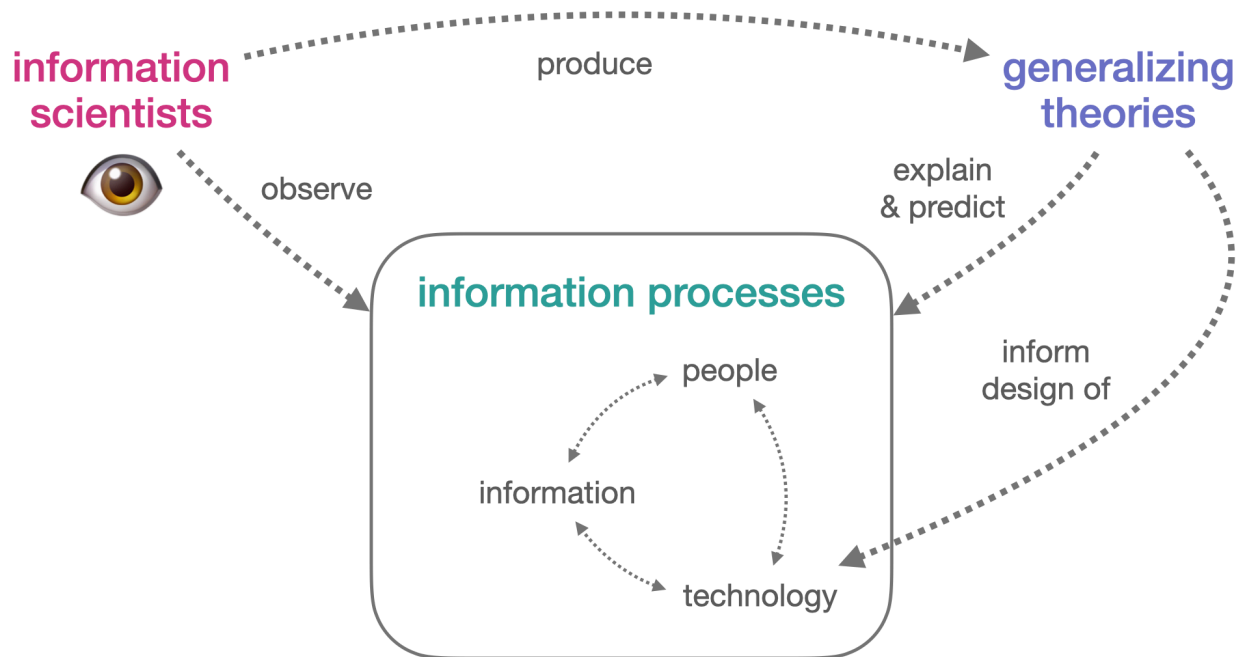
Note. An instance of "the god trick of seeing everything from nowhere" (Haraway, 1988, p. 581).

In competition with the standard paradigm of information science is a second paradigm, the critical paradigm of information studies (Figure 2). The critical paradigm posits that there are various ongoing processes of domination involving people, technology, and power. Critical scholars of information studies observe those processes, paying special attention to how power organizes the understanding, deployment, and use of information technologies. According to the critical paradigm, participants in these processes are not conscious of how their thinking is shaped by the information institutions that structure their lives (Boltanski, 2011, Chapter 2). Critical scholars, by virtue of their training and analytic tools, have developed a consciousness that others lack. This consciousness enables them to formulate critical theories that reveal the truth about the social conditions of information. Critical information scholars hope that their theories will open people's eyes, motivating changes to and eventually emancipation from the information institutions that dominate our lives. According to the critical paradigm, information studies map a substrate of power and domination invisible to those who design and use information technologies. Thus, the critical paradigm of information studies shares with the standard paradigm of information science a fundamental asymmetry between people immersed in creating, understanding, and using information, and experts who have distanced themselves from those processes and so can see more clearly.

Figure 2

*The Critical Paradigm of Information Studies*
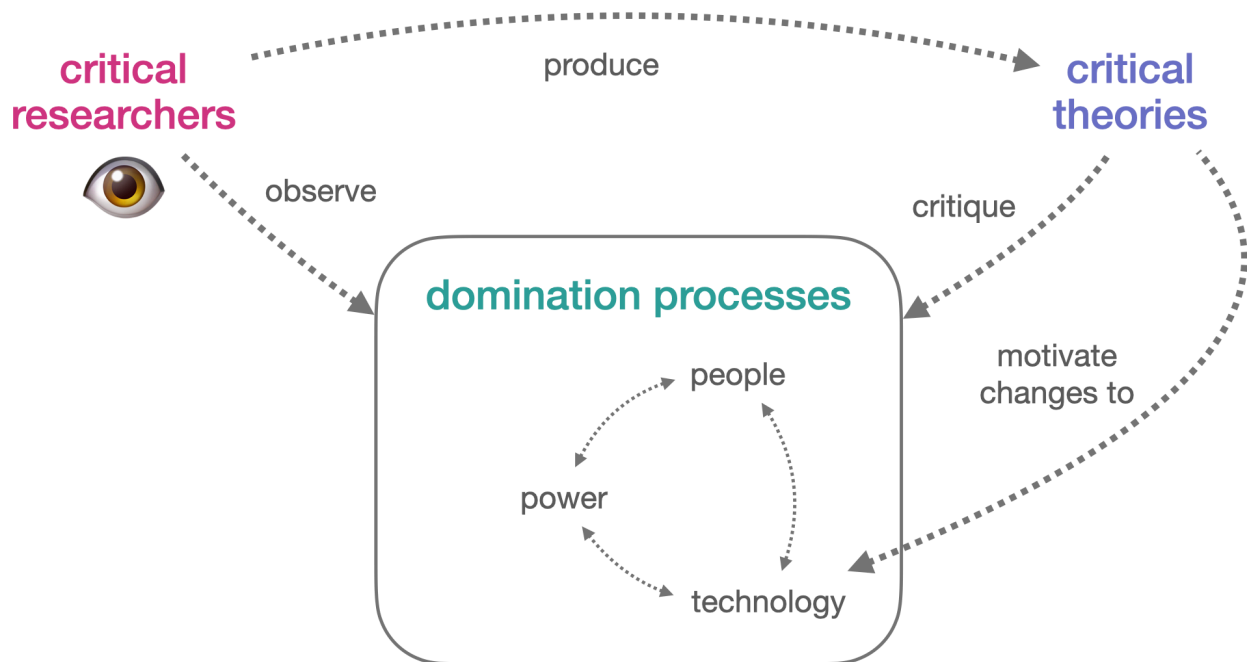
Note. A "deep asymmetry between deluded actors and the clear-minded sociologist" (Boltanski, 2013, p. 44).

Suominen and Carnap's accounts can serve as the basis for a third paradigm, one which does not share this asymmetry. This symmetric paradigm (Figure 3) supposes that all people who communicate about and by means of documents develop some communicative competence with respect to those documents. They participate in ongoing discursive practices that are mediated by documents—let's call these discursive practices "conversations." People who participate in some conversation develop communicative competence with respect to the kinds of documents that mediate the conversation. A familiarity with the history of a given conversation—a discursive tradition—is conducive to further developing one's communicative competence. Some participants in the conversation may begin participating in a second-order conversation about the discursive tradition, either because they develop a special interest in that history, or out of a desire to help others further develop their communicative competence, or both. Often such people choose to become "information professionals" of various stripes—though the second-order conversation is certainly not limited to such professionals.

Figure 3

A Symmetric Paradigm of Research-as-Practice
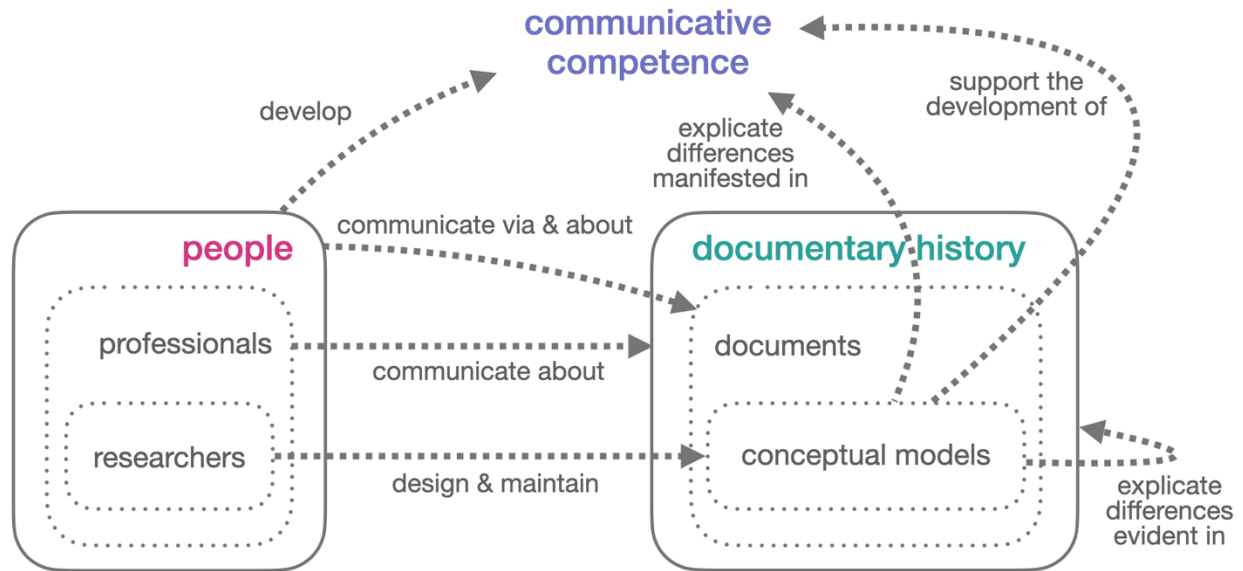
Note. People communicate via and about documents and develop different degrees of communicative competence. Some people are professionals who specialize in communicating about documentary history. Some of those professionals are researchers who design and maintain conceptual models. Conceptual models explicate differences that are 1) evident in documentary history and 2) recognized by those who have gained communicative competence, and by doing so they support the further development of communicative competence. Based on Suominen (1997, 2016).

Participants in a second-order conversation come to see continuities and discontinuities in the discursive tradition of which they are a part. They develop a sense of which distinctions have been stable over time, which distinctions seem clear at present but were not always so, and which distinctions seemed clear in the past but not now. They also become familiar with the history of attempts to explicate those distinctions for those who are less familiar, and they have opinions on how those attempts fared. This is rather abstract, so let's consider a specific example. Julia Evans is a software developer who has become well known for her "zines" that explain fundamental programming concepts (Evans, 2021a). Evans clearly considers herself to a participant in a programming culture or community—a first-order conversation. Her curiosity and enthusiasm for that conversation have led her to participate in a second-order conversation about the history of programming as a discursive tradition—for example, the history of how programmers try to

explain things to other programmers (Evans, 2021b). It is her participation in that second-order conversation that makes Evans such an effective helper of other programmers (Evans, 2017).

Among those participating in a second-order conversation about the history of some discursive tradition, there may be some who begin participating in a third-order conversation: a conversation about the conversation about the conversation. A third-order conversation is about how best to conduct a second-order conversation: what are the distinctions that matter? How should they be named and defined? How are they changing over time? Just as a second-order conversation is a reflexive specialization of a first-order conversation rather than something separate from it, a third order conversation is a reflexive specialization of a second-order conversation. A second-order conversation involves talking with and educating others about the history of a discursive tradition, and about a system of differences made evident through that history. A third-order conversation is a reflection on that second-order conversation, and it is about how to *explicate* that system of differences. In other words, it is about how to design and maintain conceptual models that support the development of competence to participate in the first-order and second-order conversations.

Conceptual modeling is what constitutes research in the symmetric paradigm. Just as professional practice is a sustained reflection on, and continuous with, "ordinary" communication, research is a sustained reflection on and continuous with professional practice. Research-as-conceptual-modeling is about attempting to communicate with others about something of mutual interest, reflecting on one's own situation and communicative practice and past failures to communicate, and designing tools—languages—to support improved communication in the future. In the symmetric paradigm researchers are engaged in the same conversation as the people they study. Researchers do not see more clearly than others on account of an external viewpoint or a clearing of deceptive fog. But they might be expected to communicate more effectively than others, thanks to a more finely honed communicative competence. And if their research is successful, then the conceptual models that they develop will help others communicate more effectively as well.

The symmetric paradigm of research-as-practice is a better description of what IS/S researchers do than either the standard paradigm or the critical paradigm. Despite the rhetoric of decades past, IS/S researchers do not discover generalized laws of information. Nor does it make sense to conceive of IS/S as a purely critical endeavor, freed from the technical work of modelling and building. The truths that IS/S research produces are partial and provisional, limited to particular contexts. They are models: useful tools for thinking and acting with at specific times and places. They do not accumulate or agglomerate into grand theories or revolutionary intellectual systems. Good IS/S research produces "good enough" descriptions of the world, capable of catalyzing ongoing processes of communication (Mazzocchi, 2008). IS/S research iteratively sustains a plurality of discursive traditions; it does not asymptotically approach essential truths. For IS/S

researchers, adopting the symmetric paradigm would mean recognizing and accepting what IS/S research already is in practice, rather than radical changes to that practice.

Nevertheless, recognition and acceptance of the symmetric paradigm would undoubtedly change the field of IS/S. Researchers would no longer be able to credibly claim expertise about "information" while lacking any specific subject-matter expertise. Advanced training in IS/S practice and research would only come after significant training and experience in some other discursive tradition such as astronomy or architecture or activism. As a result, the field would likely become even more diffuse, a network of research-practitioners spread across various kinds of organizations, rather than being concentrated in schools and departments of research universities. Furthermore, recognizing the constructive nature of explicative research—Barthes' "structuralist activity"—would mean greater engagement with the neglected aesthetic and humanistic aspects of IS/S: the arts of language and rhetoric and other techniques for dealing with symbolic media, the analysis and appreciation of ideas, devices for grasping and appreciating aesthetic structures, and historiography. A conception of IS/S research as conceptual modeling and language design closes the gap between social and humanistic study on the one side of our field and engineering research and development on the other (Wilson, 1996, p. 322).

Discussion

Treating conceptual modeling as language design is productive for at least four reasons. First, it enables seeing the different traditions of conceptual modeling as having explored different but overlapping trajectories in a single broad design space. Knowledge organization, domain modeling, and both symbolic and statistical knowledge representation can all be viewed as varieties of language design, opening the door to potentially beneficial new hybrids. For example, Carus and Ogilvie (2009), inspired by "micro-histories" of village life in early modern Europe, advocate a method of conceptual modeling that aims to bridge "the strict dichotomy between qualitative, hermeneutic understanding and quantitative, explicit theory" (p. 902). They describe this method as the iterative, dialectic construction of a meta-language that is faithful to the immanent categories of situated knowledge, while still fruitful for broader generalization and comparison. The method is distinguished from typical approaches to quantitative analysis in that

> … quantitative tests of the hermeneutic understanding do not then simply supersede that original understanding. The progress from qualitative to quantitative is not a one-way street. The qualitative, intuitive, holistic understanding derived from immersion in documents is not displaced by its quantitative explication or cross-checking in the meta-language. The quantitative cross-check is the beginning, rather, of an open-ended dialectical or mutual feedback process between the original hermeneutic understanding and quantitative data generated to test it. (Carus and Ogilvie, 2009, p. 903)

Their detailed description of the method blends elements of Suominen's vision of a historically aware KO with statistical data modeling—a possible exemplar for hermeneutic data science (Shaw, 2015).

Second, by insisting on the historical and structural dimensions of language, Suominen identifies an aspirational ideal for conceptual modeling. Agre (1995) points out that a subject classification "may embody a cataloger's understanding of [academic literatures'] social history, but it will provide little explicit representation of that history" (p. 226). As a result, subject classifications and other KO tools are far less useful for exploring those literatures than they could be. Designing better conceptual models of academic literature would require abandoning pretensions to neutrality and explicating "some frequently contested matters, such as who founded the literature, which research groups are dominant, which survey articles are definitive, which systems of ideas prefigured which others, and so forth" (Agre, 1995, p. 226). That is an accurate description of Suominen's ideal of conceptual modeling as explicating differences at the level of documentary history and structure, not only at the level of subject matter. Furthermore, Agre and Suominen's (and Carnap's) ideal of conceptual modeling firmly rejects the idea that there is clear and meaningful distinction between knowing some subject matter and knowing the structure of discourse about that subject matter (Bates, 1999). Because explication begins from vague and informal concepts, there is no way of determining in advance whether the explication of some concept will require working at the level of subject matter, or at the level of the history and structure of the discourse about that subject matter, or both (Suominen, 1997, p. 179; Carus, 2007, p. 284).

Third, Carnap's ideal of revisionary conceptual engineering offers a constructive alternative to the critical paradigm of information studies. As explained above, that paradigm typically assumes a "deep asymmetry between deluded actors and the clear-minded sociologist" (Boltanski, 2013, p. 44). Conceptual modeling offers instead a symmetric paradigm of research as sustained reflection, continuous with practice. That paradigm would risk becoming closed and self-confirming if conceptual modeling were conceived of as a purely descriptive project aimed at initiating novices into a discursive tradition. This is why Carnap's emphasis on conceptual modeling in the revisionary mode is important. Conceptual modeling should not be focused solely on describing discursive traditions but should try to intervene in and improve them as well. Growing interest in such ameliorative projects has been one of the factors driving renewed attention to Carnap's work in recent years (Yap, 2010; Haslanger, 2020). Like historically aware conceptual modeling, conceptual modeling in the revisionary mode is still largely an aspirational ideal, but there are some examples, such as the work of Susan Brown (2020) and her colleagues on the Canadian Writing Research Collaboratory ontologies. Such work points the way to a conceptual modeling practice that can make critique effective by joining it to the technical work of modeling and building (Shaw, 2019).

Finally, contemplation of revisionary conceptual modeling leads to the recognition of a grand challenge for conceptual modeling research and practice: the development of conceptual models that traverse boundaries between discursive traditions. Revisionary conceptual modeling seeks to define better concepts, but according to whose definition of *better*? Different discursive traditions have different standards for evaluating conceptual innovations, and participants in one tradition often find the standards of another tradition not only disagreeable but unintelligible (Millgram, 2015, Chapter 2). Yet many of our most challenging problems—climate change, global inequality, pandemic response—require coordination and cooperative across these different discursive traditions. And it is precisely to enable such coordination and cooperation that enhanced communicative competence is most needed. With such problems in mind, Carus (2007)—extrapolating from Carnap's ideal of explication—imagines a "language-engineering convention" promoting

> … a form of discourse that makes communication possible within some specific context where a mutual need for institution-building is acknowledged. The model for this convention would be the (idealised) social process of explication … we make ourselves understood to each other however we can, and from those tentative footholds we agree on rules of communicative interaction so that we can make ourselves better and more clearly understood for the purpose agreed on. (p. 303)

A utopian vision to be sure, but surely more pragmatic than the fantasy that a neutral "marketplace of ideas" alone can provide the necessary coordination.

Conclusion

Conceptual modeling is a broad practice encompassing knowledge organization, domain modeling, and knowledge representation. It is best understood not as a scientific process of discovery, but as a constructive process of language design. This constructive process involves both explicating differences of meaning implicit in some discursive tradition and revising those differences in order to improve that tradition. Understood this way, conceptual modeling can serve as the basis for a paradigm of research and practice that does not reproduce fundamental asymmetries between researchers and the people they study or between practitioners and the people they serve. Progress within this paradigm will involve combining methods from what have up to now been different traditions or modes of conceptual modeling. It will involve closer attention to genealogy—the historical and structural dimensions of discursive traditions. And it will involve reimagining critique as something that can go hand-in-hand with craftwork (Agre, 1997, p. 155).

The ultimate challenge for the conceptual modeling paradigm is to facilitate the building of institutions that can coordinate across different discursive traditions. Taking this challenge

seriously will require abandoning dreams of universal languages or totally comprehensive models. In practice, it might look less like engineering and more like the messy business of diplomacy, negotiating conditions for piecemeal compromises that may or may not be accepted (Stengers, 2011, Book VII). Work in this paradigm would draw on many forms of expertise that information practitioners and researchers already have, but that expertise would be applied differently. The goal would not be to design more efficient systems "to get the right information from the right source to the right client at the right time" (Mason, 1990, p. 125). Nor would the goal be to merely demonstrate once again how existing information institutions shape our thought and action. Instead, researcher-practitioners would seek to effectively communicate with others about information of mutual interest, always keeping in mind their own situation and past failures to communicate, while trying to design imperfect languages—conceptual models, always provisional and fragile—that might support improved communication in the future.

---

[1] Suominen attributes the concept of explication to Itkonen (1978), who attributes it to Pap (1958), who was greatly influenced by Carnap when they were colleagues at the University of Chicago.

References

Agre, P. E. (1995). Institutional circuitry: Thinking about the forms and uses of information. *Information Technology and Libraries*, 14(4), 225–230.

Agre, P. E. (1997). Toward a critical technical practice: Lessons learned in trying to reform AI. In G. C. Bowker, L. Gasser, S. L. Star, & B. Turner (Eds.), *Bridging the great divide: Social science, technical systems, and cooperative work* (pp. 131–158). Erlbaum.

Amrami, A., & Goldberg, Y. (2018). Word sense induction with neural biLM and symmetric patterns. In E. Riloff, D. Chiang, J. Hockenmaier, & J. Tsujii (Eds.), *Proceedings of the 2018 conference on empirical methods in natural language processing* (pp. 4860–4867). Association for Computational Linguistics. https://doi.org/10.18653/v1/D18-1523

Barthes, R. *Critical essays.* Northwestern University Press.

Bates, M. J. (1999). The invisible substrate of information science. *Journal of the American Society for Information Science*, 50(12), 1043–1050. https://doi.org/10.1002/(SICI)1097-4571(1999)50:12<1043::AID-ASI1>3.0.CO;2-X

Bengio, Y., Lecun, Y., & Hinton, G. (2021). Deep learning for AI. *Communications of the ACM*, 64(7), 58–65. https://doi.org/10.1145/3448250

Boltanski, L. (2011). *On critique: A sociology of emancipation.* Polity Press.

Boltanski, L. (2013). A journey through French-style critique. In P. Du Gay & G. Morgan (Eds.), *New spirits of capitalism? Crises, justification, and dynamics* (pp. 43–59). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199595341.003.0002

Borko, H. (1968). Information science: What is it? *American Documentation*, 19(1), 3–5. https://doi.org/10.1002/asi.5090190103

Brown, S. (2020). Categorically provisional. *PMLA/Publications of the Modern Language Association of America*, 135(1), 165–174. https://doi.org/10.1632/pmla.2020.135.1.165

Carnap, R. (1950). *Logical foundations of probability.* University of Chicago Press.

Carus, A. W. (2007). *Carnap and twentieth-century thought: Explication as enlightenment.* Cambridge University Press. https://doi.org/10.1017/CBO9780511487132

Carus, A. W., & Ogilvie, S. (2009). Turning qualitative into quantitative evidence: A well-used method made explicit. *The Economic History Review*, 62(4), 893–925. https://doi.org/10.1111/j.1468-0289.2009.00486.x

Cazden, C. B. (2011). Dell Hymes's construct of "communicative competence." *Anthropology & Education Quarterly*, 42(4), 364–369. https://doi.org/10.1111/j.1548-1492.2011.01144.x

Davis, R., Shrobe, H., & Szolovits, P. (1993). What is a knowledge representation? *AI Magazine*, 14(1), 17–33. https://doi.org/10.1609/aimag.v14i1.1029

Donahue, E. (2021). *Android linguistics: How machines do things with words* (Publication No. 28320703) [Doctoral dissertation, Duke University]. ProQuest Dissertations Publishing.

Dumontier, M., & Hoehndorf, R. (2010). Realism for scientific ontologies. In A. Galton & R. Mizoguchi (Eds.), *Proceedings of the international conference on formal ontology in information systems (FOIS 2010)* (pp. 387–399). IOS Press. https://doi.org/10.3233/978-1-60750-535-8-387

Evans, E. (2004). *Domain-driven design: Tackling complexity in the heart of software.* Addison-Wesley.

Evans, J. (2017, September). How to answer questions in a helpful way. https://jvns.ca/blog/answer-questions-well/

Evans, J. (2021a). About wizard zines. *Wizard zines.* https://wizardzines.com/about/

Evans, J. (2021b, August). Patterns in confusing explanations. https://jvns.ca/blog/confusing-explanations/

Furner, J. (2012). FRSAD and the ontology of subjects of works. *Cataloging & Classification Quarterly*, 50(5–7), 494–516. https://doi.org/10.1080/01639374.2012.681269

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems 27 (NIPS 2014)*. Curran Associates. https://papers.nips.cc/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html

Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing? *International Journal of Human-Computer Studies*, 43(5–6), 907–928. https://doi.org/10.1006/ijhc.1995.1081

Haraway, D. (1988). Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies*, 14(3), 575. https://doi.org/10.2307/3178066

Haslanger, S. (2020). Going on, not in the same way. In A. Burgess, H. Cappelen, & D. Plunkett (Eds.), *Conceptual engineering and conceptual ethics* (pp. 230–260). Oxford University Press. https://doi.org/10.1093/oso/9780198801856.003.0012

Hearst, M. A. (1992). Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 14th conference on computational linguistics - Volume 2* (pp. 539–545). Association for Computational Linguistics. https://doi.org/10.3115/992133.992154

Hjørland, B. (2007). Semantics and knowledge organization. *Annual Review of Information Science and Technology*, 41, 367–405. https://doi.org/10.1002/aris.2007.1440410115

Hymes, D. H. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics: Selected readings* (pp. 269–293). Penguin.

Itkonen, E. (1978). *Grammatical theory and metascience: A critical investigation into the methodological and philosophical foundations of 'autonomous' linguistics.* John Benjamins.

Leitgeb, H., & Carus, A. (2021). Rudolf Carnap. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy.* https://plato.stanford.edu/archives/sum2021/entries/carnap/

Lyytinen, K. (1987). Two views of information modeling. *Information & Management*, 12(1), 9–19. https://doi.org/10.1016/0378-7206(87)90068-1

Mason, R. O. (1990). What is an information professional? *Journal of Education for Library and Information Science*, 31(2), 122. https://doi.org/10.2307/40323396

Mazzocchi, S. (2008, March 18). Interoperability by friction. *Stefano's Linotype.* http://www.betaversion.org/~stefano/linotype/news/143/

Millgram, E. (2015). *The Great Endarkenment: Philosophy for an age of hyperspecialization.* Oxford University Press.

Pap, A. (1958). *Semantics and necessary truth.* Yale University Press.

Shane, J. (2021, July 2). The art of asking nicely. *AI Weirdness.* https://www.aiweirdness.com/the-art-of-asking-nicely/

Shaw, R. (2015). Big data and reality. *Big Data & Society*, 2(2). https://doi.org/10.1177/2053951715608877

Shaw, R. (2019). The missing profession: Towards an institution of critical technical practice. In *Proceedings of CoLIS, the tenth international conference on conceptions of library and information science. Information Research*, 24(4). http://InformationR.net/ir/24-4/colis/colis1904.html

Simsion, G. (2007). *Data modeling theory and practice.* Technics Publications.

Smith, B. (2004). Beyond concepts: Ontology as reality representation. In A. Varzi & L. Vieu (Eds.), *Proceedings of the international conference on formal ontology in information systems (FOIS 2004)* (pp. 73–84). IOS Press.

Stalnaker, R. C. (1967). Events, periods, and institutions in historians 'language. *History and Theory*, 6(2), 159–179. https://doi.org/10.2307/2504359

Stengers, I. (2011). *Cosmopolitics II.* University of Minnesota Press.

Suominen, V. (1997). *Filling empty space: A treatise on semiotic structures in information retrieval, in documentation, and in related research.* University of Oulu.

Suominen, V. (2016). *About and on behalf of* scriptum est*: The literary, bibliographic, and educational rationality* sui generis *of the library and librarianship on the top of what literature has produced.* University of Oulu. http://urn.fi/urn:isbn:9789526212302

Wilson, P. (1968). *Two kinds of power: An essay on bibliographical control.* University of California Press.

Wilson, P. (1996). The future of research in our field. In J. Olaisen, E. Munch-Petersen, & P. Wilson (Eds.), *Information science: From the development of the discipline to social interaction* (pp. 319– 323). Scandinavian University Press.

Yap, A. (2010). Feminism and Carnap's principle of tolerance. *Hypatia*, 25(2), 437–454. https://doi.org/10.1111/j.1527-2001.2009.01080.x