

Timothy W. Elfenbein. Metadata Pragmatics: Toward A Unified Semiotic Framework. A Master's Paper for the M.S. in I.S. degree. April, 2013. 89 pages. Advisor: Ryan Shaw

With the increasing focus on data sharing in the sciences and information organization in networked information-spaces, metadata has become a prominent area of research and activity. Scholars in library and information science and science studies have distinct approaches to describing and understanding metadata. This thesis reviews the accounts of metadata given by these two fields and then takes preliminary steps toward a unified analytic framework for research, based in Peircean semiotics as it has been developed within semiotic anthropology.

Headings:

Metadata

Signs

Semiotics theory (Communication)

METADATA PRAGMATICS:
TOWARD A UNIFIED SEMIOTIC FRAMEWORK

by
Timothy W. Elfenbein

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Science.

Chapel Hill, North Carolina
April 2013

Approved by

Ryan Shaw

For Dawn, the data to my meta

The transmission of signals is an inadequate metaphor for the interpretation of signs. “Communication” presents itself as an easy solution to intractable human troubles: language, finitude, plurality. Why others do not use words as I do or do not feel or see the world as I do is a problem not just in adjusting the transmission and reception of messages, but in orchestrating collective being, in making space in the world for each other. Whatever “communication” might mean, it is more fundamentally a political and ethical problem than a semantic one.

——John Durham Peters, *Speaking into the Air*

Contents

List of Tables and Figures.....	4
Data, Metadata, and Twenty-First Century Science	5
Purpose of Thesis.....	9
<i>Research Design</i>	14
<i>Outline of Thesis</i>	18
Semiotics.....	20
<i>Semiotics versus Semiology</i>	21
<i>Semiosis and the Third</i>	23
<i>Signs, Objects, Grounds</i>	28
<i>From Semiotics to Semiotic Anthropology</i>	31
Science Studies Accounts of Data and Metadata.....	33
<i>Socio-Technical Situations of Metadata</i>	35
<i>Metadata as Communicative Common Grounds</i>	37
<i>Data Inscription and Mobility</i>	39
Library and Information Science Accounts of Metadata.....	46
<i>Metadata as Signification</i>	47
<i>Definitions of Metadata</i>	52
<i>Metadata as Function</i>	62
Conclusion	70
References.....	71

List of Tables and Figures

Table 1. Key Peircean Distinctions.....	29
Table 5. Tensions in Long-Term Infrastructure Development	37
Table 2. Displacement in Cataloging and Metadata	51
Table 3. Dublin Core Metadata Functions and Types	63
Table 4. IFLA Metadata Functions and Types	64
Figure 1. The Third.....	25
Figure 2. Joint-Attention.....	25
Figure 3. Interaction.....	26
Figure 4. Chain of Semiosis.....	27
Figure 5. Object as Correspondence-Preserving Projection	30
Figure 6. Metadata Generation as a Third	56
Figure 7. Metadata Use as a Third	57
Figure 8. Semiotic Object 1: Stereotypical Object	58
Figure 9. Semiotic Object 2: Objectification by Participant Activity.....	60
Figure 10. Semiotic Object 3: Objectification of Text	61
Figure 11. Metadata as Instrument Realized by an Action.....	67
Figure 12. Metadata as Instrument Realized by an Action Realized by an Utterance.....	68

Data, Metadata, and Twenty-First Century Science

During the past several decades, work in the sciences has increasingly come to rely on computational power to sift and analyze bigger and bigger sets of data.¹ Whether massive amounts of data is generated by remote sensing instruments surveying environmental conditions, by text mining millions of Twitter messages, search engine queries, or digitized novels, or as the result of decades worth of census surveys, researchers are producing and using more data than ever before. Progress in the natural sciences is now thought to rely on the ability of research communities to make their data available in forms that enable reanalysis, and aggregation for meta-analysis, through grid computing (Atkins et al. 2003; Gray 2009). Humanities scholars, who are integrating the increasing store of digitized collections of libraries, archives, and museums into their work, are celebrating the wider dissemination of cultural resources and debating how big data is reshaping humanistic inquiry (ACLS 2006; Berry 2011; Rosenzweig 2003). Quantitative social scientists are using datasets from government and university-based data repositories, and more recently from social media companies, and qualitative researchers are investigating data sharing, reuse, and reanalysis (Bergman and Eberle 2005; boyd and Crawford 2012; Fabian 2008; Manovich 2012). In an effort to enable researchers to find and access, use and re-use data at larger scales, an array of projects are building out the

¹ This thesis treats “data” as a singular mass noun as it is often used among practicing scientists. See Vertesi and Dourish (2011, 534 n. 2) for the use of “data” as a collective noun, in conformity with its usage among their informants.

information infrastructure of repositories, computational networks, modeling and simulation software, and data management standards and tools.

Metadata standards development is an area of intense focus in these infrastructure projects across the sciences. Metadata—commonly understood as “data about data”—is a critical piece of information infrastructure designed to support the discovery, access, use, management, and preservation of data, documents, and other information objects. The term only came into wide use with the expansion of database management systems, networked computing, and the emergence of the Internet. Within the library community, metadata was initially seen as a method of providing organizational structure to the profusion of information resources appearing on the Internet so that they could be more effectively searched and retrieved. As metadata schemes such as the Dublin Core, Text-Encoding Initiative (TEI), and Encoded Archival Description (EAD) were developed and adopted, and as libraries pushed their open public access catalogs (OPACs) onto the open web, many library and information scientists began to understand traditional cataloging and indexing activities as species of metadata generation. The metadata schemes emerging in the sciences such as the Darwin Core, Ecology Metadata Language (EML), and Crystallographic Information Files (CIF) were less all-encompassing, pursuing domain- or discipline-specific standards and tools for particular research communities seeking more effective methods of aggregating and sharing data among colleagues.

Even as researchers express their enthusiasm for the promised benefits of easily shared data, information infrastructure projects have not achieved even rates of success across all scientific domains. While fields such as astronomy, genomics, and physics have been relatively successful at creating the culture and infrastructure for data sharing,

with data repositories such as GenBank, Dryad, and the Astrophysics Data System, seamless sharing is still not a reality (Campbell et al. 2002). Research communities that lack uniform data types or research methods, or with short histories of data sharing, face significant difficulties in their standardization efforts (Bowker 2000), and researchers have varied reasons for, and concerns about, sharing their data (Borgman 2012; Carlson and Anderson 2007). The process of infrastructure development often surfaces organizational misalignments, divergent incentives for participation, and intra- and inter-disciplinary tensions between differently-positioned actors and agencies. The difficulty of implementing even well-designed infrastructure projects suggests a need to better understand the dynamics of socio-technical systems (Hughes 1989; Star and Ruhleder 1996).

For science studies scholars—those who study the life of the institutions and practices of science—data-intensive research has become a productive site of inquiry, especially for those looking at information infrastructure (Bowker et al. 2010; Edwards et al. 2009; Ribes and Lee 2010). For library and information scientists, the focus of much research into scientific metadata is oriented toward finding methods of increasing interoperability between proliferating metadata schemes through the use of such tools as crosswalks and application profiles (Baker et al. 2005; Willis, Greenberg, and White 2012). These two approaches to metadata, one studying the place of metadata practices in evolving socio-technical formations, the other focused on standards development and the improvement of information architecture, often speak past one another. Recent research initiatives, however, from science studies scholars within information science programs

have started to bridge the gap between the concerns and vocabularies of these research communities.²

² See Edwards et al. (2007) for the report on a National Science Foundation (NSF) workshop that tackled both practical issues of infrastructure development and theoretical issues of establishing a research agenda for infrastructure studies.

Purpose of Thesis

This thesis takes preliminary steps toward a unified analytic framework with which to approach metadata. The framework is analytic in that it seeks to identify and distinguish the constituents and fundamental relations of the phenomena to be studied. It is unified in that the proposed analytic covers the phenomena discussed in both the library and information science and science studies literatures, including the conceptual, linguistic, and social aspects of metadata.³ Analytic frameworks are necessary in the initial stages of social science research and one is adopted either before the collection of empirical data or in the first steps of analysis.⁴ The work of elaborating a framework relies both on previous empirical research and the critical reflections of higher-level theories and philosophies of the human sciences. The framework forwarded in this thesis is built on the assumption that the constituents and relations of metadata are species of social phenomena in general and therefore amenable to social, cultural, and linguistic analysis. Established analytic approaches to empirical research on metadata are reviewed, not with the aim of critically evaluation but to suggest a different place to initiate research in the future. The purpose of the framework is to provide a metalanguage for the description of metadata and its situational contexts, with the hope that new descriptions will lead to new insights.

³ See Søren Brier (2006; 2008) for his proposal of a unified pan-semiotic framework for information science.

⁴ See Charles Ragin (1994) on the place of analytic frames alongside theories, data, and synthesizing images in social research.

The approach to metadata proposed in this thesis emerges out of the insights of two recent works. Paul Edwards, Matthew Mayernik, Archer Batcheller, Geoffrey Bowker, and Christine Borgman (2011), in an article synthesizing a variety of ethnographic research on scientific metadata initiatives, wish to create a more capacious understanding of metadata by including within it all communication underlying data sharing, what they label as “common ground.” Even as metadata is designed to reduce “data friction”—the cost in time, energy, and human attention required at the “points where data moves between people, substrates, organizations, or machines” (669)—metadata itself also has costs.⁵ Scientists often rely on ad-hoc and informal communication when sharing data to minimize the costs of producing scheme compliant metadata, prompting the authors to ask how they might develop a broader understanding of metadata as a communicative phenomenon.

D. Grant Campbell’s (2005) article proposes a novel approach to classifying the activities involved in generating metadata, using modes of signification as the means to identify and distinguish classes. Whereas most efforts to understand metadata within library and information science focus on the various functions metadata supports, he analyzes metadata’s specialized use of language through structuralist literary theory. By appealing to modes of signification such as metaphor and metonymy Campbell develops a framework that focuses on how metadata does its work, as opposed to the ultimate purpose of that work.

⁵ Both “data friction” and “metadata friction” are metaphors originally articulated by Paul Edwards (2010) in his book on the information infrastructure of climate science.

This thesis proposes a framework with which to understand the metadata as communicative common ground, approached as a form of signification. To do so, it reviews the ways library and information science and science studies describe and understand metadata. The research in these two literatures provides different ways of conceptualizing and studying metadata, and identifies different aspects of metadata—as contextual information, organizing structures, practices, processes, and institutions. This thesis seeks to add to the insights of this research by appealing to semiotics and anthropology. The semiotics of Charles Sanders Peirce, as interpreted through semiotic and linguistic anthropology, provides the analytic foundations for a new approach to metadata.

A number of scholars in library and information science have appealed to versions of semiotics in their work.⁶ Michael Buckland (1997), D. Grant Campbell (2005), Douglas Raber and John Budd (2003), Richard Smiraglia (2001), and Julian Warner (1990) use the semiology of Ferdinand de Saussure and his acolytes in structuralist literary theory, while others such as D. C. Blair (1990), Søren Brier (2006; 2008), Jens-Erik Mai (2001), and Torkild Thellefsen (2002) apply Peircean semiotics to questions of knowledge organization and representation.⁷ Science studies scholars have also incorporated various aspects of semiotics into their work, although for most of the field

⁶ This thesis distinguishes between two traditions within the field by using “semiology” to denote the tradition descending from the work of Ferdinand de Saussure, and “semiotics” to refer to that of the work of Peirce. When the distinction is less crucial, “semiotics” is used to denote the study of sign phenomena in general.

⁷ Blair (1990) reviews semiotics so as to set it up as a foil to his preferred Wittgensteinian approach to the problem of linguistic representation in information retrieval systems. In his review, Peircean semiotics is intermixed with Saussurean semiology, as was prevalent in most representations of the field until fairly recently. The influence of Umberto Eco’s (1979; 1986) synthetic presentation of semiotics contributed to this intermixing, as is the case in Blair’s work.

semiology is the reigning model (Høstaker 2005; Lenoir 1994; Watts 2008). This thesis looks to Peircean semiotics as it has been interpreted and developed within anthropology so as to approach metadata as constitutive of and a constituent within everyday practices in social worlds. Just as contemporary anthropological scholarship seeks to “bring theories of the sign into a full, robust articulation with accounts of human actions, self-consciousness, and social power” (Keane 2003a, 410), this thesis attempts to bring metadata within the orbit of such accounts.

The aim of this thesis is to bring to light aspects of metadata that are less visible in existing research by using a different metalanguage to describe its constituents and processes. It is inspired by the work of scholars such as Bernd Frohmann (2004), and Dilip Gaonkar and Elizabeth Povinelli (2003). Frohmann presents a unified non-cognitivist framework for describing both the practices of science and scientific information systems: “My intent is . . . to use recent studies of materiality and labour processes involved in scientific work as resources for reorienting studies of scientific information toward scientific documentation” (17). Extending a Wittgensteinian critique of meaning and using practice-based science studies, Frohmann sets forth a program for a materialist and processual account of documentation:

“Specific directions for research in scientific documentation and documentation generally are proposed: (1) translating talk about ideas, concepts, and information into talk about occasioned utterances and inscriptions; (2) interpreting inscriptions as actants, that is, as among the elements exhibiting agency and resistance in scientific labour; (3) expanding the range of scientific activities in which documents might play a role beyond legitimate, authorized science; (4) expanding the range of documentary circuits that figure in scientific labour; and (5) recuperating the role of history in documentation studies. (21–22)

For Frohmann, properly attending to the consequences of Wittgenstein's later philosophy requires a fundamental rethinking of the analytic framework used to understand both scientific practice and the documentation systems that support it.

Gaonkar and Povinelli (2003) identify a similar analytical shift in their outline of a coalescing research program in the interdisciplinary journal, *Public Culture*. The journal was established in 1988 by scholars interested in the role played by flows of cultural forms, such as media, tourism, and sports, in the process of globalization. While the goal was to understand cultural forms in motion, authors writing for the journal often fell back on established humanistic methods to produce readings of the social meaning of texts. Gaonkar and Povinelli recognized a shift in approach of newer submissions to the journal and with them the emergence of a genuine "ethnography of forms" (391, emphasis removed):

"First, [the essays] offer form-sensitive analyses of public texts, events, and practices that do not succumb to the temptation of reading for meaning. Second, they foreground the cultures of circulation and transfiguration within which those texts, events, and practices become palpable and are recognized as such. Third, they disclose the play of supplementarity that enframes and ruptures the enterprise of public recognition whatever its object." (386)

The shift promoted by Frohmann, Gaonkar, and Povinelli is toward modes of analysis that attend to the materiality, movement, and transformations that underlie the constitution of particular social worlds. This type of analysis still listens for actors' understandings, for instance, the ways they talk about the meanings of scientific articles or media events. The work of analysis, however, is to describe the material forms and affordances, patterns of circulation, and the work of discourse itself in characterizing and solidifying objects, that underlies actors' accounts. If forms are privileged above actors' understandings in this type of analysis, and this style of analysis thereby seems to be a

later-day species of structuralism, the new focus on materiality and circulation point to methods through which the ethnography of forms pushes past the *aporias* of earlier analytic frameworks. The attempt might be described as post-Foucauldian in its effort to find a new path forward:

[Foucault] has sought to avoid the structuralist analysis which eliminates notions of meaning altogether and substitutes a formal model of human behavior as rule-governed transformations of meaningless elements; to avoid the phenomenological project of tracing all meaning back to the meaning-giving activity of an autonomous, transcendental subject; and finally to avoid the attempt of commentary to read off the implicit meaning of social practices as well as the hermeneutic unearthing of a different and deeper meaning of which social actors are only dimly aware. (Dreyfus and Rabinow 1983, xxiii–xxxiv)

A similar shift is proposed by this thesis in describing the basis of common ground, what is usually understood as semantic relationships in the library and information science literature. The goal is to make visible and describe the pragmatics of metadata as it is created, circulated, and inserted into new settings. Such a project requires the articulation of a different analytic framework than those underlying existing research—a metalanguage with which to describe the objects, processes, norms, and discourse of metadata.

Research Design

This thesis was designed from the outset as productively tentative and speculative: as an exercise in abductive analysis (Locke 2010; Timmermans and Tavory 2012). It began with a hunch, that the present understanding of metadata does not fully account for the semiotic labor invested by metadata generators and users. This sense emerged out of readings of research such as Alenka Šauperl's (1999) exhaustive qualitative study of the process of subject determination by library catalogers: "People tend to associate like

things together. . . . What ‘like’ is depends on the person’s previous experience. People’s memories get stored in mental schemas. . . . It seems reasonable to believe that catalogers try to activate their memory, their schemas, with their search for associated records” (252). So goes the summary of social cognition and categorization, and its use to illuminate the cataloging process. This is as theoretically daring as Šauperl allows herself to be in an otherwise carefully observed study that is more comfortable flying low to the ground. This kind of theoretical conservatism in qualitative research, where extensive observational data results in a behaviorist model and a few low-level theories, is a consequence of the empiricist methodological tendencies of much work in library and information science.⁸ While perhaps not unexpected for a field still anxious over its status as a social science, empiricist qualitative research such as this leaves one with the feeling that as much is being missed as caught in the analysis for want of better conceptual tools.⁹

Paired with this initial hunch about what might be missing in the analysis of metadata was another about what might be discovered. This hunch emerged as an speculative analogy: perhaps the relationship between data and metadata is similar to that between language and metalanguage: “The semiotic–functional approach [to reflexivity in language] moves beyond a concern with the relation of linguistic reference to knowledge and considers language use as a form of social action, most particularly, as communicative action. . . . Metacommunicative and metalinguistic activity takes place all the time to help structure ongoing linguistic activity” (Lucy 1993, 18). Typical examples

⁸ For a less jaundiced view of modeling from within information behavior research, see Case (2007).

⁹ See Mai (2001) for a more theoretically adventurous analysis of subject indexing.

of this activity include language about the structure or functioning of language (“The word is a regular verb with an ‘-ed’ added to the end indicating past tense”) and reports of speech (“So I said, ‘That is wicked cool,’ and Ralph was like, ‘I know, right’”). Both these examples also contain subtler forms of reflexivity that index the event of speaking and speakers: the temporality of a verb in the past tense is reckoned by the event when the verb is spoken; and the denotation of the pronoun “I” is alternatively determined by the reporting speaker and the use of quotations to index the speaker of the reported speech. The analysis of metalanguage has produced a number of seminal insights into the ways language and communication work. Perhaps a semiotic-functional approach could open those aspects of metadata that are unapproachable from more empiricist research. Perhaps such an analysis could produce insights on the constituents and relations of communicative activity comparable with those of the linguists and anthropologists who developed it.

The hunches that led to this thesis were mediated through and generated by a particular intellectual background. Locke (2010) argues that “an observation’s status as ‘fact’ or as ‘surprising,’ and as the impetus for abduction, is never given purely; it is always mediated by modes of perception, by background perspectives and theories” (2). Timmermans and Tavory (2012) go further, stating that abductive analysis “rests for a large part on the scope . . . of the theoretical background a researcher brings to research. Unanticipated and surprising observations are strategic in the sense that they depend on a theoretically sensitized observer who recognizes their potential relevance” (197). Common to these understandings of abduction analysis is the cultural and biographical situatedness of the researcher that conditions assumptions, expectations, and purposes.

Whereas in other research traditions this would be seen as a potential source of unconscious bias, in abductive analysis situatedness is a source of innovation and insight.¹⁰ A propensity to look for certain kinds of conceptual gaps in analysis, reinforced through the training of intellectual habits, underlies a cultivated sense of surprise (Brubaker 1993). In abductive analysis, there is no way to understand the researcher as incidental to the path an analysis will take.

Timmermans and Tavory (2012) are not, however, advocating for a mode of sociological research where a researcher's favorite theory dictates the course theory building will take (Tavory and Timmermans 2009; cf. Burawoy 1998). They argue for an approach that is grounded first and foremost in abductive reasoning, but which relies on the methodological heuristics of grounded theory: "The methodological precepts of grounded theory can stimulate abductive reasoning through a process of revisiting, defamiliarizing, alternative casing in light of theoretical knowledge. The surprise, puzzle, or anomaly that may trigger a novel theory then emerges methodologically through careful data analysis against a background of cultivated theoretical expertise" (Timmermans and Tavory, 2012, 180).

The design of this thesis seeks to propose the contours of an alternative casing, in part by revisiting the research on metadata. Although Timmermans and Tavory's (2012) discussion of methodological heuristics to promote abductive reasoning is oriented to ethnographic research, its suggestions are also pertinent for analyses of existing literature.

¹⁰ Hans-Georg Gadamer (1985 [1965]) argues that understanding is fundamentally embedded within language and culture, an insight that forms the foundation of hermeneutic analysis. For important discussions of the implications in hermeneutic social science, see Bernstein (1983), Taylor (1985), and Rabinow and Sullivan (1987).

“Researchers should think through different conceptual and theoretical frameworks in both coding and memo writing. . . . Each casing abstracts and highlights different aspects of the phenomenon, rendering it comparable to different phenomena and turning it into a generalization that then can be linked to other fields and theories” (177). This thesis seeks to propose a new casing for metadata, one different from those proposed by library and information scientists and science studies scholars. It adapts an analytic framework and concepts from semiotic anthropology to approach metadata through a different casing. The hope is that relatively unexplored aspects of metadata will become more visible with the proposed alternative casing. The purpose is not to generalize over a sample of cases, but rather to articulate an analytic framework from which the interpretations and generalization of empirical cases might later emerge.

Outline of Thesis

The rest of the thesis proceeds as follows: a brief survey of semiotics is given, starting with Saussurean semiology so as to distinguish it from Peircean semiotics, and then of the fundamental constituents of Peirce’s model of semiosis. The model is presented through the interpretations and elaborations of contemporary linguistic and semiotic anthropologists, most especially Paul Kockelman. The implication of a pragmatic analytic framework for the study of metadata, and sign phenomena more generally, closes this section of the thesis.

After the survey of semiotics, a broad overview of science studies is presented before looking at the subfield of information infrastructure studies. Several ethnographic and qualitative studies of metadata and other information infrastructure initiatives that

take a sociological approach are reviewed. The article by Edwards et al. on the communicative common grounds of metadata is further explored before turning to the concepts of inscription and immutable mobiles from actor-network theory. The section closes with an initial analysis of the pragmatic work underlying the decontextualization of metadata from its originating contexts.

In the final section, the library and information science literature on metadata is reviewed starting with D. Grant Campbell's semiological approach to classifying metadata generation activity, which holds signification as central to the understanding of metadata. Influential definitions of metadata are reviewed, followed by preliminary redefinitions using the semiotic framework. A review of research that understands metadata through its functions follows, and the section ends with a semiotic analysis of metadata understood as an instrument.

Semiotics

A unified analytic framework for metadata must be able to account for the many ways the research literature understands metadata, which will be reviewed below: as surrogates or supplements of designated objects, instruments for resource-related functions, arrangements of labor and circulating inscriptions, and as means of communicating with potential future users. Semiotics has an integrated set of analytical definitions, distinctions, modes of relation, and a range of theories that provide the foundations for such a framework. A brief survey of semiotics is presented below, focusing on those parts that provide conceptual tools relevant to the framework.

Semiotics, which is usually loosely defined as the science of signs, has been described as a broad topical field of research encompassing everything from biosemiotics to computational semiotics, however it is better characterized as something more than either a disciplinary field or a methodological tradition. The specific corner of the semiotic tradition that is of interest in this thesis is that identified with the American philosopher, Charles Sanders Peirce (1839–1914), and later refined by linguists and anthropologists, most prominently Roman Jakobson (1990) and Michael Silverstein (1976; 1993). Anthropologists since the 1980s have sought to draw tighter links between cultural and linguistic anthropology, accomplishing this in part by reinterpreting Peirce's semiotic framework to address language use in interaction (Agha 2007a; Hanks 1996), the interactional emergence of textuality and the social lives of texts (Bauman and Briggs 1990; Hanks 1989; Keane 1997; Silverstein and Urban 1996), and the place of ideologies

and other reflexive aspects of language in everyday use and in scholarly analysis (Keane 2003a, 2003b; Lee 1997; Lucy 1993; Mertz and Parmentier 1985; Parmentier 1994).

Their work seeks to account for the constitution and inter-relations of words, things, actions, subjects, and social formations within an analytic framework of semiotic analysis (Manning 2003; Hankins 2013; Kockelman 2005, 2013a).¹¹ Central to the reinterpretation of Peirce's semiotics is the reevaluation of the pervasive legacy of Ferdinand de Saussure's semiology in the humanities and social science.

Semiotics versus Semiology

Peirce's model of semiotics has gained greater influence in the past twenty years as scholars sought ways of addressing central analytic divisions within social, linguistic, and literary theories based in Saussure's semiology. Saussure, whose *Course in General Linguistics* (1966 [1916]) established the analytic framework of twentieth-century linguistics, posits a distinction between *langue* (language) and *parole* (speech), or abstract linguistic sign system over against its concrete instantiations in speech. Language was approached as a relational system governed by syntactic rules and semantic relations, and this complete structure was presupposed in the analysis of materially- and historically-situated utterances of speakers: "From the very outset we must put both feet on the ground of language [*langue*] and use language as the norm of all other manifestations of speech [*parole*]" (Saussure 1966, 9, emphasis removed). This

¹¹ See Mertz (2007) for a review of the various emphases and directions of semiotic anthropology since the 1970s. See Duranti (2003) for the longer history of the relationship between linguistic and cultural anthropology in the United States.

distinction enabled linguists to concentrate their research on stable synchronic language structures.

The split of *langue* from *parole* complicated reigning substantive models of the referential relationship bounding meaning to words. For Saussure, language should not be approached as collections of words paired with meanings or concepts. Concepts are determined not by their referents but rather their differential position within the total structure of *langue*. Saussure's model of the sign distinguishes between the sensory sound-image of speech, called the *signifier*, and the psychological concept, the *signified*. While the relational structure of language determines the meaning of signifieds, the ties of signifiers to signifieds is weakened, bound only by arbitrary convention. The principle of the arbitrary nature of the sign, along with the separation of the material signifiers from conceptual signifieds, was broadly influential in structuralist anthropology, linguistics, literary theory, and social theory.

One of the consequences of Saussure's influence has been the increasing division between materialist and idealist accounts of social phenomena. The "radical separation of the denotational sign (qua sign) from the material world" (Irvine 1989, 248) has hindered the emergence of analysis that effectively addresses the full panoply of social life. As Webb Keane (2003b) argues, "efforts to bring theories of the sign into a full, robust articulation with accounts of human actions, self-consciousness, and social power are still commonly hampered by certain assumptions built into the lineage that runs from Saussure to post-structuralism" (410). By bringing the assumptions of the Saussurean legacy to the fore, this strand of semiotic anthropology has attempted to develop a fuller

account of social life by returning to, and arguing from, the perspective of Peirce's model of semiotics.¹²

Of the scholars advancing this line of research, Paul Kockelman's (2013a) work is one of the more thoroughgoing attempts to build a unified semiotic framework in view of the critiques of Saussurean understandings of sign phenomena.¹³ Kockelman's body of work is unusually systematic, detailed, and granular, aiming to articulate precise definitions and principled distinctions into a unified analytic. The work enables analysis to approach meaning as a linguistic, logical, and social process. His analytic framework begins with an essay (2005) that

enumerates, defines, and interrelates key semiotic terms. In part, this is done to provide a theory of semiosis and a metalanguage for doing semiotics; and, in part, this is done to argue against certain pervasive and erroneous assumptions about signs. While broadly Peircean in its framing, the point of this essay is neither to expound nor to espouse Peirce; rather, the point is to use his work as a starting off point to develop a theory of semiosis that can illuminate that ensemble of processes that usually fall under the headings of language, culture, and mind. (233)

The summary of semiotics below predominantly follows from Kockelman's work.

Semiosis and the Third

Whereas Saussure forwarded a model of signification based on the arbitrary relationship between signifier and signified, Peirce proposed a more complex trichotomous model of

¹² See Agha (2007b) for a discussion of the legacy of Saussure in the discipline of linguistics. He argues that the epistemic project of linguistics was narrowed by its reception of Saussure, while other parts of the human sciences have retained a broader understand of the object of linguistic study.

¹³ The recently published monograph by Kockelman (2013a) provides a synthesis of his analytical and theoretical work. This thesis relies on earlier iterations of the same material (2005, 2006a, 2006b) as well as work published after the monograph (Kockelman and Bernstein 2013; Kockelman 2013b).

semiosis, “any process involving the relation between a sign, an object, and an interpretant” (Kockelman 2005, 234).¹⁴ Peirce (1955b) sets out the inter-relationship of a third thusly: “A sign, or *representamen*, is something which stands to somebody for something in some respect or capacity. It addresses somebody, that is, creates in the mind of that person an equivalent sign, or perhaps a more developed sign. That sign which it creates I call the interpretant of the first sign. The sign stands for something, its object” (99).¹⁵ Although Peirce has usually been read as requiring human cognition or mental interpretants, later theorists have generalized the semiotic framework to encompass non-cognitive semiosis (Watts 2008, Kockelman 2005). Kockelman (2005) reformulates the definition of a third to better reflect this generalization and to clarify, as much possible, Peirce’s circuitous prose: “A sign stands for its object on the one hand, and its interpretant on the other, in such a way as to make the interpretant stand in relation to the object corresponding to it[s] own relation to the object” (234, emphasis removed). In Kockelman’s generalized definition, a third involves a complex of standing-for relations:

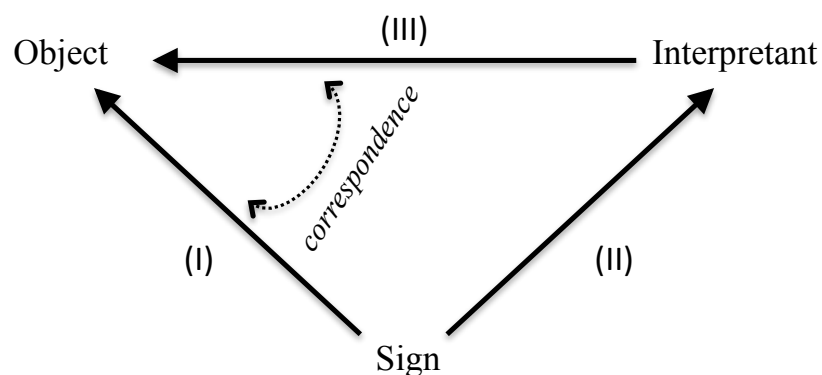
- (I) the sign stands for an object;
- (II) the sign stands for its interpretant;
- (III) the interpretant stands for the object such that the relation between sign and object corresponds with the relation between interpretant and object.

This understanding of the third distinguishes itself from other interpretation of Peirce by focusing on the relation of correspondence between two relations, the last part of the complex (figure 1).

¹⁴ Peirce used the term “sign” to denote both the triadic relation between sign, object, and interpretant, and the first constituent of that relation. To reduce confusion, the triadic relation will be called a “third,” following Kockelman (2005, 236).

¹⁵ While Peirce uses “sign” as a subset of the category “*representamen*” (Hankin 2013, 217n1), to avoid confusion, only the term “sign” will be used for the first constituent of a third.

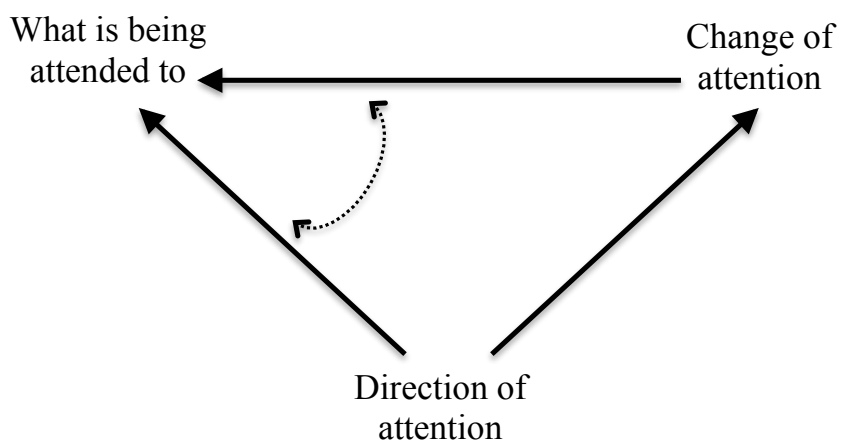
Figure 1. The Third.



Source: Adapted from Kockelman (2005, 236), figure 1.

Kockelman (2005, 237–39) describes a number of types of thirds as examples of semiosis. An instance of joint-attention, where Alice turns her attention to what Bob is attending to, is an third involving Bob's direction of attention (sign), what is being attended to (object), and Alice's change of attention (interpretant). In this type of third, which provides a useful model of intersubjectivity, the object is the focus (figure 2).

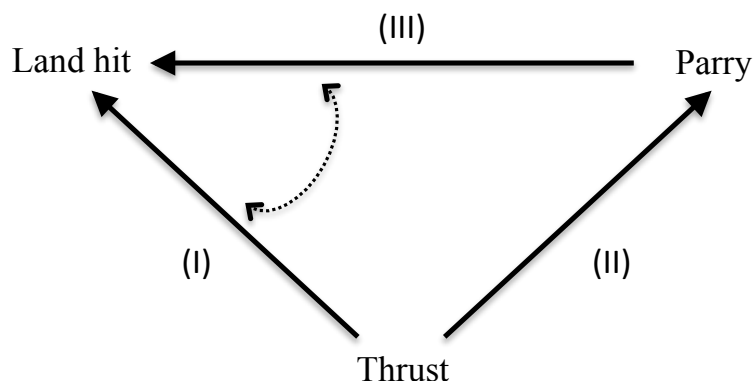
Figure 2. Joint-Attention.



An interaction involving controlled behavior is a third. For example, in a fencing match, Carol's thrust (sign) is met with Dennis's parry (interpretant) to defend against the rapier point landing a hit (object). Notice that the object in this example is not an object

stereotypically understood—the rapier or Dennis’s body—but rather the purpose or second part of the controlled behavior, landing a hit. Restating this example to parallel the formal definition of a third: (I) Carol’s thrust stands for the rapier’s impact on the one hand, and (II) Dennis’s parry on the other, (III) in such a way as to make the parry stand in relation to the impact corresponding to the thrust’s relation to the impact (figure 3). Generalizing the definition of this type of third, an interaction consists of a controlled behavior (sign), a purpose (object), and a reaction based on the recognition of the purpose (interpretant). Other types of thirds identified by Kockelman include, pair-part structures, affordances, instruments, roles, third-party relations, and commodities,¹⁶ and other types, such as Sebeok’s (2001) symptoms, could be added.

Figure 3. Interaction (Fencing example).

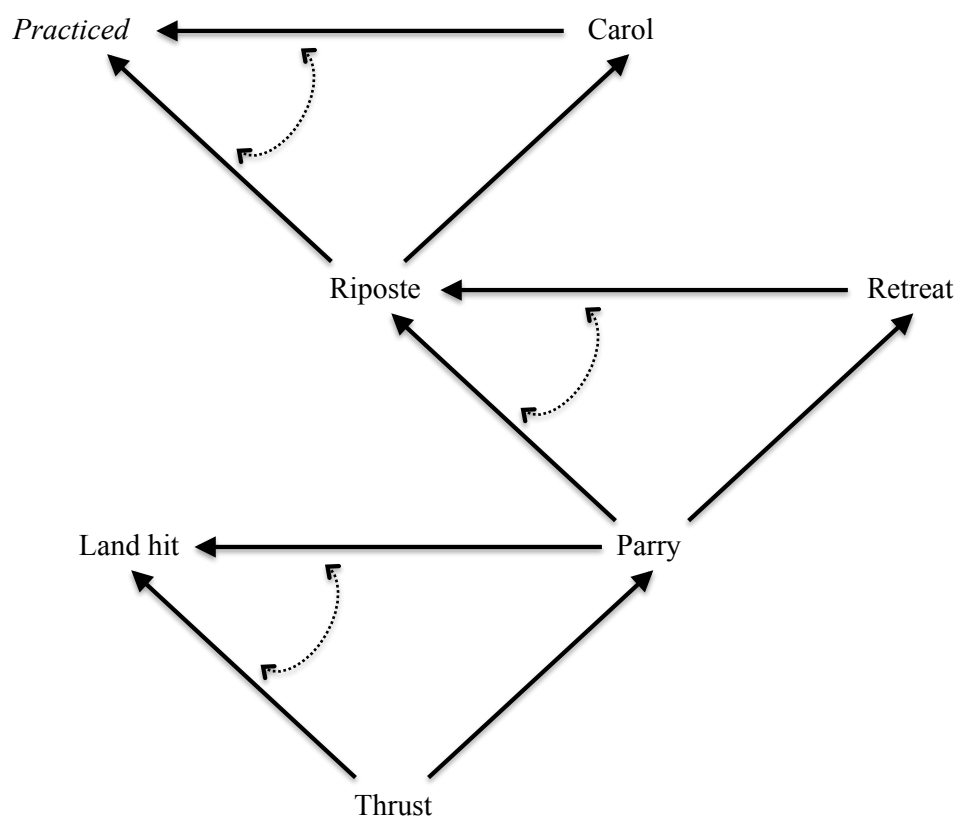


An important aspect of Peirce’s model of sign phenomena is the unbounded, open-ended nature of semiosis. Any component of one third might simultaneously or subsequently be a component of a different third. Continuing with the fencing example, competitor Dennis’s parry could become a sign, with Carol’s hasty retreat (interpretant)

¹⁶ See Kockelman (2006a) for a full reanalysis of Marx’s model of the commodity based in Peirce’s trichotomy of sign–object–interpretant instead of Hegel’s dichotomy of subject–object.

the defense against Dennis's riposte (object). Here, the purpose remains the same (to land a hit), even if the body targeted shifts from one competitor to another, continuing the interaction. A component of a third might also become a component in a different type of third. Subsequent to the interaction above, Dennis's riposte (sign) is taken as an indication by Carol (interpretant) that Dennis is well practiced (object) (figure 4). This last third comes closer to the usual understanding of semiosis as a model of inference requiring cognition. As the previous examples attest, however, cognition is but one possible effect of sign functioning.

Figure 4. Chain of Semiosis (Fencing example).



Signs, Objects, Grounds

Peirce's understanding of what could function as a sign is expansive, including "pictures, symptoms, words, sentences, books, libraries, signals, orders of command, microscopes, legislative representatives, musical concertos, performances of these" (MS 634:17–18, quoted in Parmentier 1994, 3).¹⁷ The common aspect of these disparate phenomena, what enables them to function as signs that stand for something else, is their perceptibility by the senses. Signs are typically understood to be sensible, segmentable, stable, persistent, symmetrically sensible to signer and interpreter, and compositional, although Kockelman (2005, 240–41) argues that these attributes do not hold for all signs. He points to embodied signs such as social or intentional statuses that, under the right semiotic framing, could be signs. While these attributes do not necessarily obtain in any particular case, Peirce identifies a relational requirement of all signs: "A sign does not function as a sign unless it be understood as a sign" (MS 599:32, quoted in Parmentier 1994, 4). Nothing acts as a sign unless it is taken as a sign by an interpretant, or, from the opposite direction, the sign must effect an interpretant.

Peirce distinguishes three kinds of signs: qualitative signs (*qualisigns*), singular signs (*sinisigns*), and legislative signs (*legisigns*). A qualisign is a "mere quality" that could possibly be paired with an object. A sinisign is an "actual existent" quality that is paired with an object in space and time, or a token. A legisign is a "general law," a quality that must be paired with a class of objects, or a type (quotes from Peirce 1955b, 101). For example, red is a possible qualisign. To become a sinisign, red must be

¹⁷ Citations to the collection of Peirce's manuscript held at Harvard University use the cataloging system organized by Robin (1967).

embodied in an actually existing form, say, in a particular apple. Red is but one qualisign bundled together among other possible qualisigns that might be embodied in sensible forms. In an apple, red is contingently bundled together with qualisigns such as hardness or spherical shape.¹⁸ Red might also be a general type or legisign, say, Pantone color number 19-1664, “True Red,” used to match instances of red paint. In this typology of signs, two types of sinisigns can be distinguished: sinisigns (tokens) that fall under a legisign (types) are called replicas by Peirce (1955b, 102), and unique or unprecedented sinisigns are called singularities by Kockelman (2005, 241).

In Peirce’s analytic framework, legisigns presupposed sinisigns, and sinisigns presupposed qualisigns, or, from the other direction, qualisigns are embedded in sinisigns, and sinisigns in legisigns. This relationship of presupposition or embedding is a key trait in Peirce’s framework based in his phenomenological categories of firstness, secondness, and thirdness (1955a). Firstness is pure qualitative feeling or possibility and is the mode of signs and qualisigns. Secondness is the unmediated actuality of brute fact, the mode of objects and sinisigns. Thirdness is mediated necessity or law, the mode of interpretants and legisigns (Watts 2008) (table 1).

Table 1. Key Peircean Distinctions.

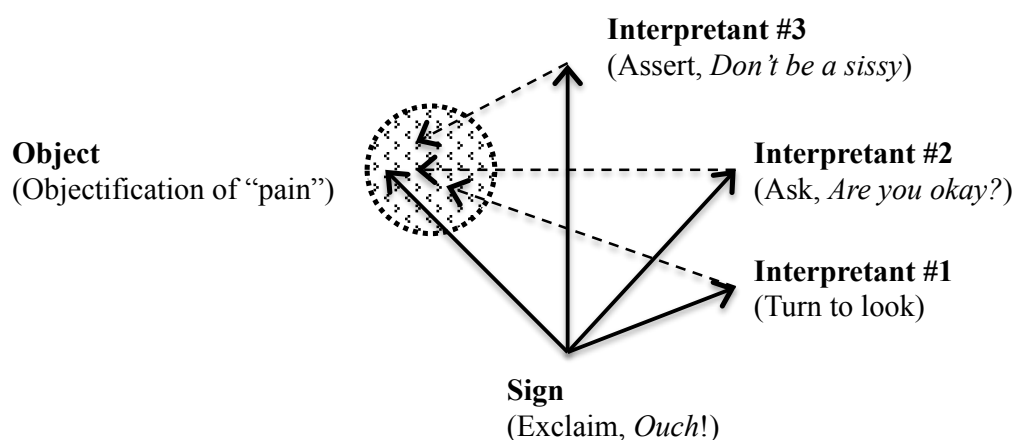
	Third	Sign	Ground
Firstness	Sign	Qualisign	Iconic
Secondness	Object	Sinisign (token)	Indexical
Thirdness	Interpretant	Legisign (type)	Symbolic

Source: Adapted from Kockelman (2005, 235), table 1.

¹⁸ On the importance of the bundling of qualities, see Keane (2003b, 214–15).

Kockelman (2005), reformulating Peirce’s prose again, generalizes the concept of the object by defining it thusly: “the object (of a sign) is that to which all (appropriate and effective) interpretants (of that sign) correspondingly relate” (242, emphasis removed). This definition treats the object as a “correspondence-preserving projection” (ibid.), a relatively abstract focal point of all interpretants of a sign, which highlights the situational context (appropriateness) and pragmatic efficacy (effectiveness) of interpretants. In Kockelman’s example, if someone experiences pain and in reaction yells “Ouch!,” any appropriate and effective interpretant would need to relate to the same object to which the sign (“Ouch!”) relates (figure 5).

Figure 5. Object as Correspondence-Preserving Projection (Pain example).



Source: Kockelman (2005, 242), figure 2.

Peirce attends to the relations between each set of constituents, but the sign–object relation, called grounds, receive what has become the most influential treatment. The typology of three grounds includes: iconic relations based on similarity of qualities or resemblance; indexical relations based on spatio-temporal or causal contiguity; and symbolic relations based on convention. As with the constituents of a third and the sign types, grounds belong with Peirce’s phenomenological categories of firstness,

secondness, and thirdness. Likewise, symbols presuppose indexes, which in turn presuppose icons (table 1). Examples of iconic thirds include portraits, diagrams, or any other mimetic images. A classic example of an indexical third is smoke, which is caused by, and spatio-temporality contiguous with, fire. In the illustrations above, the relation of an object of attention to a person's gaze (taken as a sign) is an index, as is the relation of a thrust to a hit in the fencing match. If Carol had read Dennis's riposte as indicating that he had been trained in the classical style, this would be an example of a symbolic third.

From Semiotics to Semiotic Anthropology

The brief introduction of semiotics above provides few hints at the implications of adopting the analytic framework of semiotic anthropology for analyzing metadata. Others have used various forms of semiotic analysis within library and information science and science studies: what does Peircean semiotics as developed within anthropology contribute to these analyses? Semiotic anthropology's move away from the analysis of linguistic signs, and signification more generally, that privileges semantics has to far reaching consequences for analyses seeking to integrate signification and society.

Michael Silverstein (1976), in his seminal essay "Shifter, Linguistic Categories, and Cultural Description," argues that the semantic analysis of traditional linguistics was overwhelmingly concerned with only one purposeful aspect of language, the referential or denotational function. A broader understanding of language use, however, would see the referential function as only one among a number of functions. By placing semantics and semantic analysis within this broader, multifunctional understanding of pragmatic meaning and pragmatic analysis, Silverstein provides a means of studying language in

society: “Once we realize that distinct pragmatic meanings yield distinct analyses of utterances, we can sever our dependence on reference as the controlling functional mode of speech . . . We can then concentrate on the manifold social pragmatics that are common to language and every other form of socially constituted communication in society” (20–21).¹⁹ Kockelman (2005) follows Silverstein by arguing for the primacy of pragmatics, with semantics as a subordinate category: “Pragmatics treats the meaning of linguistic signs in terms of their appropriateness in context and their effectiveness on context. Semantics treats that subset of pragmatics in which the meaning of utterances is inferentially articulated—or, equivalently, in which propositional and/or conceptual content is conferred upon the objects of signs” (245).

By framing semantics and reference as embedded within normative and pragmatic relations of appropriateness and effectiveness, semiotic anthropology opens into view semiotic infrastructures not visible in most library and information science and science studies research. Such an analytic framework provides new ways of understanding metadata as communicative common ground and as one process for enclosing meaning (Kockelman 2013b).

¹⁹ Silverstein’s mentor, Roman Jakobson, initiated the analysis of semantics within a pragmatic frame by identifying such linguistic phenomena as shifters (1971b [1957]) and iconicity in grammar (1971a [1966]). It was Silverstein, however, who first worked out the significance of these phenomena for socio-cultural analysis. See Caton (1987) for a summary of Roman Jakobson’s importance in anthropology. See Lucy (1993) for the important distinction between logical–linguistic and semiotic–functional approaches to reflexive language.

Science Studies Accounts of Data and Metadata

One of the unexpected benefits—at least for those studying metadata—of the growing prominence of data sharing in the sciences has been the new salience of science studies. The field offers a range of conceptual tools and approaches for describing and understanding science as epistemological, technical, and practical endeavor.

The past half-century has seen the proliferation of sociological and historical studies of the institutions and practices of science and the emergence of a variety of subfields and research traditions. Intellectual historians attribute the energy driving the expansion of the field to the debates following in the wake of Thomas Kuhn's *The Structure of Scientific Revolutions* (1996 [1962]), which broke from earlier whiggish accounts of the history of science (Hess 1998; van House 2004, 8).²⁰ Nancy van House (2004) argues for the salience of the interdisciplinary subfield of science and technology studies (STS) for information science, in particular for research interested in the construction and communication of scientific knowledge. Within information science, bibliometrics and scientometrics have well-established connections to the early structural-functional sociology of science developed by Robert Merton in the 1930s. Van House (2004) focuses her attention, however, on more recent qualitative approaches in STS that “critically examines the nature of knowledge and especially the collective

²⁰ For a review of earlier research in science studies, see Turner (2008).

processes and practices of knowledge production, interpretation, and use in technoscience” (5)

The field site for early qualitative research in STS was the scientific laboratory, studied using ethnographic methods (e.g., Knorr-Cetina 1981; Latour and Woolgar 1979; Lynch 1985). These studies distinguished themselves from earlier approaches by taking contemporary, everyday practices of science as their object of study, and by relying on ethnographic observations instead of scientists’ reports on their own actions. This helped to reveal the messy material and social practices—the “bricolage, tinkering, discourse, tacit knowledge, and situation actions” (Fujimura 1992, 170)—on which scientific knowledge is built. Despite much misunderstanding of these laboratory ethnographies, especially by practicing scientists, the purpose of this approach is to not to explain away knowledge. As Janet Vertesi and Paul Dourish (2011) stress, “going behind the sciences in the laboratory . . . is not a question of airing dirty laundry, but is aimed at analyzing the *how* of science in order to better understand its role in society, and to design policies and systems that best fit and support how science is done” (534).

Qualitative research in STS during the past thirty years has expanded its attention beyond the scientific laboratory to large socio-technical infrastructure systems (Edwards et al. 2009; Hughes 1989; Star and Ruhleder 1996). A subset of the field focuses on information infrastructure in a range of settings, such as the development of information and communication technologies (Bowker et al. 2010; Ribes and Lee 2010), health informatics systems such as electronic health records (Coopmans 2006; Jirotko et al. 2005; Ure et al. 2009), and cyberinfrastructure for research in a range of earth sciences (Edwards 2010; Faniel and Jacobsen 2010; Borgman, Wallis, and Enyedy 2007). Much

of this literature touches on the place, movement, and translations of data within the institutions and practices studied.

Socio-Technical Situations of Metadata

One approach taken by science studies scholars to metadata initiatives describes the social or political aspects of infrastructure initiatives using ethnographic and other qualitative methods. Florence Millerand and Geoffrey Bowker (2009) study the development and implementation of the Ecological Metadata Language (EML) by the Long-Term Ecological Research Network (LTER), an association of over 2300 scientists working at twenty-six research stations spread throughout the United States and Antarctica. Their study focuses on the centralized standards-development team and dispersed station information-managers, relating their different perspectives on the metadata initiative. Millerand and Bowker use two analytic concepts, enactment and trajectory, to organize the study: enactment distinguishes between the material components of a technology and the perceptions and practices of the people bringing the technology to life in particular contexts; and trajectory highlights the narrative temporalities used by different groups to describe the same technology project. The authors show how the social roles and status of groups within the LTER change, and how the EML project contributes to a shift in the division of labor in the daily work of research, even as the metadata project struggles to realize its goals. Millerand and Bowker bring to light the usually invisible labor of infrastructure enactment and the divergent narratives of success, taking up themes from the field of computer supported cooperative work (Gerson and Star 1986; Bowker and Star 1999; Suchman 1994). They

only glancingly address the informational consequences of negotiated settlements over the meanings of metadata properties, for instance, how locally-derived data ontologies are folded into the standard.

Like Millerand and Bowker, many science studies scholars addressing the development of information infrastructure for data sharing use conflicts among groups to identify underlying institutional structures, motivations for the contribution of labor to collective projects, and temporal rhythms of development. David Ribes and Thomas Finholt's (2009) study of recurring tensions in information infrastructure development is organized around the concept of "the long now," an orientation toward institutional futures. They aggregate their observations from four case studies of projects in the earth and environmental sciences: the Geosciences Network, the LTER (described above), the Linked Environment for Atmospheric Discovery project on mesoscale meteorological phenomena, and the Water and Environmental Research Systems Network project developing a remote sensing network for hydrology. Analyzing the perspectives of participants as they go about their daily work of designing, developing, and implementing information infrastructure projects, Ribes and Finholt identify a common set of ongoing concern—points of tension where participants attempt to weigh what they see as conflicting priorities and potential outcomes. The concerns include how to motivate contributions to the project, how to align the many possible end goals of the endeavor, and how to the design of the project that could be adopted by future users. These three concerns are tracked across three different scales of infrastructure: "(I) [infrastructure] is a technological venture, seeking to deploy durable resources to support work, automate tedious tasks, and enable collaboration; (II) it is a matter of human work, organization,

and maintenance . . . ; and (III) it is an institutional venture, seeking to provide stable and accessible services to communities at national and international levels” (377–78) (table 5).

Table 5. Tensions in Long-Term Infrastructure Development, According to Concerns and Scales of Infrastructure.

	Institutionalizing	Organizing Work	Enacting Technology
Aligning End Goals	Project vs. facility	Planned vs. emergent	Inclusion vs. readiness
Motivating Contribution	Individual vs. community	Development vs. maintenance	Research vs. production quality systems
Designing for Use	Communities vs. constituencies	Research vs. development	Today’s requirements vs. tomorrow’s users

Source: Ribes and Finholt (2009).

Ribes and Finholt’s goal of gathering a variety of organizational, technical, and contextual phenomena together under one analytic framework emerges as one method of tackling the difficulty of approaching the heterogeneous terrain of socio-technical systems. It is a method that acknowledges “the ‘soft’ foundations (i.e., organizational and human) infrastructure will require just as much research and work as its ‘hard’ foundations (i.e., technical)” (391).

Metadata as Communicative Common Grounds

Much of the recent developments in science information infrastructure studies can be traced to the efforts of a small group of scholars located within information science programs. Christine Borgman at the University of California, Los Angeles, Geoffrey Bowker at the University of California, Irvine, and Paul Edwards at the University of

Michigan have each led research institutes and NSF-funded initiatives,²¹ and published extensively on information systems, metadata development, scientific practices, and the evolving institutional structures of the sciences (Borgman 2007, 2012; Borgman, Wallis, and Enyedy 2007; Bowker 2007; Bowker and Star 1999; Edwards 2010; Edwards et al. 2007). Their work, and that of their growing list of former-students, is on the leading edge of science information infrastructure studies. Their co-authored essay (Edwards et al. 2011), along with (then) graduate students Matthew Mayernik and Archer Batcheller, represents an important expansion in the imagination of what studies of data sharing and metadata practices should encompass.

Edwards et al. (2011) bring together qualitative observations of metadata initiatives from three projects of different scale: several hundred datasets of simulated runs of global atmospheric circulation models deposited in the Earth System Curator web-portal system; the ongoing collaboration between instrument engineers and environmental scientists at the Center for Embedded Networked Sensing; and the stalled metadata standardization project of the LTER (described above). The authors see data sharing as involving a general problem of human communication, the need to “establish mutual agreement, and adjusting and confirming shared understanding” (670–71). The observations of the projects are organized through a conceptual distinction between metadata as product (“static, definitive descriptions of data characteristics”) and as process (“ad hoc, incomplete, loosely structured, and mutable descriptions of data”)

²¹ Borgman leads the data-practices team of the Center for Embedded Networked Sensing (CENS) and is co-principle investigator of “Knowledge Infrastructures: Astronomy Data.” Edwards was the lead-principle investigator for the “History & Theory of Infrastructure” conference and report (Edwards et al. 2007), which Bowker was a co-principle investigator. Bowker and Edwards are co-editors of the *Infrastructures* series at MIT Press. All three were co-principle investigators for the “Monitoring, Modeling, and Memory” project.

(667). While standards developers work toward the realization of metadata-as-product, researchers' daily practices of data-inscription and sharing rely on informal communications, including emails, notes, and face-to-face discussions, to effectively share and reuse data.²² The common function of metadata as product and as process—to provide enough grounding for a shared understanding of data—could be achieved either way, even if the ad hoc method puts into question the vision of a world of frictionless data sharing.

Edwards et al. (2011) are interested in the communicative function of metadata, describing methods for establishing and maintaining the understanding necessary to reuse data: “Through metadata, those charged with making data available effectively say to prospective users, ‘Here is what you need to know about these data.’ In other words, scientific metadata provide the information necessary for investigators separated by space, time, institutions, or disciplinary norms to establish common ground” (672). Their analysis of the basis for common grounds, however, is only briefly touched on with a short discussion of the concept of “grounds” from discourse analysis (671).

Data Inscription and Mobility

Inscription has been a central analytical tool in science studies since its first articulation by Latour and Woolgar (1979) in their study of knowledge construction at the Salk Institute (Lenoir 1997). As ethnographic observers determined to develop their own understanding of laboratory science, Latour and Woolgar seized on patterned acts of

²² See Zimmerman (2007) for ecologists' methods of finding, acquiring, and validating datasets prior to the advent of networked data repositories.

inscription—the successive transformations of materials, instrument outputs, annotation markings, external papers, and typed manuscripts into an end result, new scientific articles. Reflecting later on his study, Latour (1986) notes that he was struck “by the way in which many aspects of laboratory practice could be ordered by looking not at the scientists’ brains (I was forbidden access!), at the cognitive structures (nothing special), nor at the paradigms (same for thirty years), but at the transformation of rats and chemicals into paper” (3). Inscriptions were valuable both for defining an object of empirical observation and establishing the analytic level at which Latour and Woolgar would describe the world of the laboratory. Instead of appealing to epistemological theories to explain scientific practice—as a set of philosophically-justified actions—the ethnographers would stick to the mundane activities of everyday scientific work. This stance toward the construction of scientific facts has been called anti-epistemological, “more about *logistics* than logic” (Blok and Jensen 2011, 27).

To understand how scientific facts, whether constructed in the laboratory or field, travel and accumulate, Latour (1987) appeals to the concept of immutable mobiles. The significance of inscription is its transformation of aspects of the world into forms that are stable, easy to move, gather together, compare, and combine. Latour uses the example of the eighteenth-century voyages of Lapérouse, the French explorer and cartographer. The purpose of Lapérouse’s expedition in 1785 was not to observe the Pacific coasts of the Americas and East Asia, and then return to recount his discoveries to Louis XVI: it was to produce inscriptions that traveled back to France where they could be collected, catalogued, combined, superimposed, and transformed into other texts. The expedition was successful because Lapérouse’s inscriptions reached France, even though his ship

and crew never returned. As an analytical tool, the concept of immutable mobiles is useful because it focuses on the materiality of communication and the organizational and persuasive capacities of inscriptions. Although Latour was engaged in historical analysis and primarily concerned with the qualities of inscriptions on paper, he also presciently imagined the effect computing might have on inscription and immutable mobiles: “If inventions are made that transform numbers, images and texts from all over the world into the same binary code inside computers, then indeed the handling, the combination, the mobility, the conservation and the display of the traces will all be fantastically facilitated” (228).

Although Latour devised inscription and immutable mobiles as concepts that would enable a non-cognitivist approach to studying the sciences, their usefulness is limited by a continuing analytical division between signs and materiality (Frohmann 2004; Lenoir 1994). Latour visibly struggles as he appeals to concepts from semiology while disavowing them at nearly the same moment.²³ In one work, he argues “semiotics remains inadequate because it persists in considering only texts or symbols instead of also dealing with ‘things in themselves’” (Latour 1988, 183). Yet, a few years later in an influential glossary setting out the meaning of a series of foundational terms for socio-technical studies, he again appeals to a tortured form of semiotics:

Semiotics: The study of how meaning is build, but the word “meaning” is taken in its original nontextual and nonlinguistic interpretation; how one privileged trajectory is built, out of an indefinite number of possibilities; in that sense, semiotics is the study of order building or path building and may

²³ Maxim Waldenstein (2008), before staging his own dialogue between science studies and Soviet semiotics, notes that “although Latour originally revolutionized the field of science studies by introducing the ideas and practices of ethnomethodology and semiotics, he later pronounced semiotics to be inadequate” (222).

be applied to settings, machines, bodies, and programming languages as well as texts; the word socio-semiotics is a pleonasm once it is clear that semiotics is not limited to signs; the key aspect of the semiotics of machines is its ability to move from signs to things and back. (Akrich and Latour 1992, 259)

Latour seems to recognize the potential and limit of semiological analysis and is both unable to do without it or resolve its central analytic division. Other important actor-network theorists have perpetuated the analytical division between sign and material phenomena, even as they continue to struggle to move beyond it. John Law (2009), for instance, expresses his wish to rename the family of methodological concerns organized under the label of “actor-network theory” as “material semiotics,” a phrase that includes both sides of the dichotomy instead of offering a path toward its resolution.²⁴ Even as actor-network theorists have struggled with the limitations of semiology, it should be noted that a number of anthropologists have pointed to Latour as someone working against the pervasive dualisms entrenched in Saussure-influenced social theory (Keane 2003a, 411).²⁵

Ethnographic research on health information systems provides examples of how concepts from science studies are used to understand data and metadata. Catelijne Coopmans (2006), for instance, draws lessons from a British telemedicine initiative which built a digital database of medical images to look more attentively at the “particular circumstances under which the relationships between data, digital technologies and mobility is articulated” (2):

²⁴ That Law (2009) fail to cite any semiotician, even when forwarding a new title for his own field of study, suggests that his use of the term comes with no commitment to seriously engage the semiotic literature.

²⁵ For discussions of actor-network theory from the perspective of semiotic anthropology, see Engelke (2007), Hankins (2013), Hull (2012), Manning (2012), and Watts (2008).

Data are surprisingly often assumed to automatically and necessarily become (more) mobile when they are combined with information and communications technologies (ICTs). . . . As a result, instead of asking precisely what is meant by ‘data mobility’ and how it is practically arranged, the analysis jumps straight to a discussion of effects and consequences. . . . This paper aims to indicate an alternative direction for a sociology of data mobility. . . . It does so on the basis of two simple questions: what does it mean, and what does it take for data to become mobile?” (2)

These questions are posed about a database of digital mammograms, developed by the British health service to more efficiently use radiologists and distant experts. Coopmans complicates Latour’s (1987) picture of immutable mobiles by describing the sociotechnical configuration in which both film and digital mammograms operate. Both forms of inscription are mobile, although in different ways and with differing implications for the sociotechnical configuration that supports each. She effectively criticizes the database designers’ assumption that ICTs necessarily produce more mobility.

Others, such as Berg and Goorman (1999), have also inquired into the expansion of uses of electronic medical records as data. They focus on the situational-boundedness of medical information with the context of clinical care, along with the evidentiary assessments and structures of authority between nurses and doctors. The kind of data produced in situations of clinical care can only be made useful to other possible users, such as insurance companies, epidemiologist, or hospital managers, with additional work. Berg and Goorman use their case study to propose a law of medical information: “The further information has to be able to circulate (i.e. the more different contexts it has to be usable in) the more work is required to disentangle the information from the context of its production” (52). Both these studies of medical data are pursuing questions not addressed by Latour, but his analytical tools help direct the questions nonetheless.

Berg and Goorman's law of medical information points to the work required to extract or disentangle information from its originating contexts but it says little about the arrival of information in new contexts. Coopmans touches on this process by looking at the sociotechnical configuration that supports the interoperability of medical images across contexts of production and interpretation. Studies of semantic interoperability (Howarth 2005; Zeng and Chan 2010) inquire into both sides of the process, but when semantics is understood as a subset of pragmatics, the focus of the analysis will of necessity shift. Studies of data and metadata interoperability would need to expand their purview beyond the analysis of semantic equivalents to inquire into underlying pragmatic relations that enable the communication of semantic content.

Paul Kockelman and Anya Bernstein (2013) do just this in their discussion of the semiotic technologies of temporal reckoning that determine the relative portability of meaning beyond particular contexts. Semantically, the utterances "04/17/2013," "last Wednesday," and "when Fido ran away" might be equivalent expressions for reckoning a date. What the expressions communicate across contexts, however, varies greatly depending on their indexical ties to the events of speaking and speakers, as well as the spread of the semiotic conventions used to interpret them. The expressions also communicate more than a reference to a date, for instance, differentially placing emphasis on the event of speaking, the speaker, and the event used for reckoning. Even the most portable of the expressions signifies more than a temporal referent: the syntax of the numerical string (a legisign) could index its creator's American, as opposed to British, upbringing.

Analyses of interoperability that only inquiry into semantic and syntactic relations of equivalence, to the exclusion of the pragmatic relations, assume that reference is the sole signifying function of metadata. Silverstein (1976) noticed that “out of context, we can only have recourse to the referential mode in determining the meaning of utterances, which, with certain exceptions for shifters, is essentially ‘context-free.’” (Silverstein 1976, 47). Metadata is a semiotic technology for entextualizing, contextualizing, and recontextualizing data and documents, enabling them to shed their contextual ties so they can travel far and wide.²⁶ The drive toward metadata understood as free-floating propositional information not only overlooks the pragmatic semiotic processes underlying semantic meanings, but also unknowingly smuggles in a politics of universality in the guise of pure communication.²⁷

²⁶ See Silverstein and Urban’s (1996), whose analysis of entextualization and contextualization will be discussed below. See Almklov (2008) for a case study of the decontextualization and recontextualization of data among geologists and engineers in subsurface oil exploration.

²⁷ See Chakrabarty (2000) for a postcolonial analysis of the translation of semantically equivalent terms mediated through the terms of a presumed universal language: “The claim [of Newtonian science] would be that not only do H₂O, water, and *pani* refer to the same entity or substance but that H₂O best expresses or captures the attributes, the constitutional properties, of this substance” (85). See Espeland and Stevens (1998) for a sociological analysis of commensuration—the comparison of different entities with a common metric—as a general social process.

Library and Information Science Accounts of Metadata

Scholars have worked with an array of overlapping understandings of metadata, contributing both to the terms promulgated and confusion over its use. Although the term is of recent coinage, emerging out of the database and statistics community, as a principle of information organization it has been in use for considerably longer, for instance as an approach to bibliographic control in library science (Burnett, Ng, and Park 1999; Greenberg 2010). The term came to prominence in library and information science as librarians struggled to adapt their traditional approaches to cataloging for the new and unruly information space of the World Wide Web. With the expansion of networked computing and the emergence of the Internet as the central system for information searching and dissemination, metadata standards were seen as a crucial component of information infrastructure (Weibel 1995). Although a few scholars wish to distinguish cataloging from metadata creation (e.g., Heery 1996)—basing the distinction on whether the activity concerns physical objects or digital resources, or simply whether the information resides in a library catalog or not—the trend is toward the recognition of any formal documentation and description of resources as metadata (Greenberg 2003; Mayernik 2011; Vellucci 2000).

The review of metadata definitions and typologies of metadata functions offered below is not comprehensive, but seeks to highlight important or common aspects among the work of library and information scientists. These are aspects for which a semiotic definition of metadata will need to account. The review begins with D. Grant Campbell's

semiological framework for metadata, which provides an initial step toward a semiotic reanalysis.

Metadata as Signification

D. Grant Campbell's (2005) framework for distinguishing between two general types of metadata generation activities deserves greater engagement than it has as yet received in library and information science. The novelty of Campbell's framework is its appeal to structures of signification for identifying basic activities related to metadata generation across diverse domains. The problem Campbell seeks to address with his framework, similar to many of the other scholars discussed below, is an issue of classification: how to recognize similar or distinguish dissimilar information activities across heterogeneous terrains. Metadata as a tool of information organization developed out of a number of different communities including librarians, computer scientists, text encoders, and database designers, and it is created and used not only in its traditional institutions, such as in libraries by formally trained professionals, but by a host of others in non-traditional information spaces pursuing divergent objectives. Out of all these communities, institutions, and evolving practices, how can a few general classes of metadata activity be isolated and identified?

Campbell starts by selecting two functional metadata classes, arguing that they are central if not exhaustive functions of all metadata: metadata for resource discovery and resource use. He then enumerates a few attributes common among the different metadata traditions. First, metadata is data about data, or data that represents data for a specified purpose. Second, metadata works as a specific language, usually a more structured subset

of a larger language. Third, metadata is used in information spaces that are too large to be navigated effectively by individuals without the aid of information systems: “Metadata are designed to facilitate information systems; they function as a specialized language that is similar to natural language, but employed in a unique way; and they serve the purpose of representing information and information-bearing objects” (Campbell 2005, 61).

Up to this point, Campbell’s analysis shares some similarities with Kathleen Burnett, Kwong Bor Ng, and Soyeon Park’s (1999) framework for describing and classifying metadata. Their proposed definition of metadata also seeks to encompass the two source traditions they identify, bibliographic control and data management. The two functions of metadata they identify are resource discovery/identification and use, and they argue that the characterizations of source data by metadata, or “aboutness” relationships, are the central mechanism by which metadata works. Burnett, Ng, and Park see two types of characterizations, which map onto the two functions: intrinsic characterizations support discovery whereas extrinsic ones support use. While these metadata function classes are the same as Campbell’s, and he also emphasizes the representational relationship of source data to metadata, Campbell introduces a fundamentally different analytic principle with his second common attribute.²⁸

Instead of relying on function for his classificatory principle, Campbell shifts toward an alternative principle, the range of signifying relations embedded in linguistic structure. Advancing a broad analogy between the types of signifying relations in poetic

²⁸ Campbell’s third attribute, the need for navigation in information space, is not further analyzed and can be safely assumed as a contextual prerequisite for any information system using metadata.

language and metadata, he argues that literary theory, which seeks to understand literary and poetic uses of language, could provide analytical tools for understanding metadata.²⁹

The structuralist literary theory that developed out of Saussurean linguistics could be applied to the types of specialized uses of language that occur in metadata:

Poetics, . . . like bibliographic description, constitutes a series of specialized languages that are embedded in, but distinct from, natural language, that draw attention to their own difference from natural language, and that work within larger systems of meaning that are drawn from their inter-relationships within bodies of literature, in addition to their connections to what lies outside literary systems. Structuralist theory, therefore, which examines linguistic systems as self-enclosed and self-referential structures of meaning, should have some light to shed on the practices of metadata, both in its traditional garb of bibliographic description and its newer manifestations as Dublin Core headers, TEI headers, Encoded Archival Descriptions (EADs), geospatial information, and the like. (Campbell 2005, 63)

Campbell details a number of central concepts from structuralist theory including the distinction between synchronic and diachronic approaches to language, metaphor and metonymy as modes of signification, and literary displacement. Saussure made the distinction between the diachronic or historical developments of languages and their synchronic or structural totality at any one moment in time, which can be used to understand, in Campbell's examples, the derivations of original works along a diachronic view, or the necessary synchrony of namespaces and ontologies.

More important for the purposes of this thesis are Campbell's deployment of metaphor, metonymy, and displacement. Roman Jakobson (1960) proposed the distinction between metaphor and metonymy in his seminal study of poetics. Metaphor is a connection between a tenor and a vehicle based on simultaneous identity where the

²⁹ Campbell appeals to what Lucy (1993) identifies as the "literary-performance approaches" to reflexive language.

attributes of the vehicle are taken on by the tenor. Metonymy, on the other hand, is a relationship of contiguity or association between entities. Whereas Jakobson uses Russian verse to illustrate these relationships, Campbell uses metadata. For instance, the MARC codes of a bibliographic record are the tenor that connects to the descriptive identities of the vehicle, the fields, for records to be readable. The relationship between a resource and its metadata, say an article and its abstract, is metonymic: associative but not qualitatively identical. Metadata that takes one part to stand for the whole, such as a surrogate record for a book during search, is a specific type of metonym called a synecdoche.

Displacement is an extension of the distinction between metaphor and metonymy, elaborated by Northrop Frye (1977). For Frye, “metaphor creates an identity that is both apt and yet untrue” (Campbell 2005, 65) and therefore shares the same structure as myth, which presents itself as a recounting of what happened while also asserting that it did not really happen as recounted. Displacement is the movement away from myth, where mythical structures are increasingly submerged as realist genre conventions come to the fore. Romance was one step along the way, with its metonymic conventions of mythical structure: “the structures of romance are surrogates of mythical structures, rather than direct expressions of them, displaced ‘in a human direction and yet, in contrast to “realism,” [tend] to conventionalize content in an idealized direction’ (Frye 1977, 137)” (Campbell 2005, 66). These levels of increasing displacement can be identified, Campbell argues, in different types of metadata. Descriptive metadata, MARC coding, TEI markup, and other metadata oriented to resource use are mythic or metaphorical. At one level of displacement is surrogate metadata such as bibliographic records, TEI headers, or other metadata that stand in place of the resource they describe. At the realist

end of displacement, Campbell identifies metadata that eschews their own underlying conventions, such as automatic document clustering, search engine algorithms, or the arrangement of books on a shelf. Ontologies that understand themselves as mapping the entities and relations that make up reality could be considered the furthest reach of displacement from myth (table 2).

Table 2. Displacement in Cataloging and Metadata.

Figurative Device	Literary Analogue	Cataloging Analogue	Metadata Analogue
Metaphor	Myth	MARC coding; Crosswalks	TEI version of XML schemas; Marked up documents
Metonymy	Romance	Bibliographic description	Metadata descriptions; TEI header; Dublin Core
————	Realism	Serendipitous browsing	Free-text searching; Automatic clustering

Source: Adapted from Campbell (2005, 67) table 1.

The figurative devices described by Campbell leads to his recognition of a spectrum of metadata generation activities that consist of various mixtures of metaphor and metonymy. A bibliographic record stands in a metonymic relations to its resource, a surrogate that represents and stands in for its object during the process of discovery. The standardized structure of the record, however, is metaphoric with its MARC field delimiters, coding, and controlled vocabularies for descriptions. A Dublin Core metadata header stands in a metonymic relation to its web page, whether as a harvested record in a metadata repository or as a synecdoche standing in for the whole. The HTML markup of the page's content and the XML tags surrounding the metadata values set up metaphoric connections of identity.

While Campbell's aim is to create a comprehensive classification of activities related to metadata generation, by appealing to signification as his principle of

discrimination he highlights a different way of understanding how metadata does its work. The library and information science definitions and typologies of functions reviewed below focus on aspects of resource description and use, and relatedly, on the intrinsic and extrinsic characterizations that support discovery and use. Campbell's metaphor/metonymy spectrum of signification places emphasis on the signifying relations, as opposed to aspects of the object or the consequent purpose supported, and displacement hints at the underlying signifying relations that have been suppressed or forgotten in an activity that understands itself as producing transparent descriptions and identifications.

While Campbell's creative adaption of structuralist literary theory opens new aspects of metadata to scrutiny, a semiotic framework prompts a shift beyond a semiology-based analysis. His framework places its emphasis on the types of relations between metadata and objects, or signifiers and signifieds. This method of unearthing signifying relations is a crucial step, but also limited by its semiological assumptions of the constitutive division between sign and things. A semiotic analysis of metadata shifts the focus from arbitrary relations between signifiers and signifieds to the creation of appropriate and effective projections through thirds. The triadic model of semiosis opens signification to the pragmatic and normative relations.

Definitions of Metadata

A number of scholars have attempted to identify and classify common aspects of metadata schemes across domains and institutions. Several overarching definitions have emerged from these overviews and attempts at synthesis, as well as the identification of

aspects of metadata that are often lost in their high-level definitions. Burnett, Ng, and Park (1999) provide a definition that seeks to incorporate what they identify as the two primary metadata traditions, bibliographic control and data management: “Metadata is data that characterizes source data, describes their relationships, and supports the discovery and effective use of source data” (1212). They argue for the centrality of “aboutness” in metadata, or the characterization of source data by metadata, with the bibliographic control approach focusing on representations of source data and the data management approach focusing on use.

Jane Greenberg (2003) defines metadata as “structured data about an object that supports functions associated with the designated object,” and further specifies the object as being “any entity, form, or mode for which contextual data can be recorded” (1876, emphasis removed). This definition has two important differences from that of Burnett, Ng, and Park. First, it emphasizes the functional relationship of metadata to data as opposed to relationships of characterization. For characterization to be considered metadata in Greenberg’s definition, it must be paired with a purpose such as the discovery, authentication, or preservation of the object. The focus on functional relationships highlights metadata’s orientation to the future, promoting the “specified functions surrounding the life of the designated object” (1877). Second, Greenberg’s definition incorporates a far more expansive understanding of possible objects. This definition of the object makes it possible to generalize metadata as a process that occurs in numerous situations and contexts, including those not usually addressed by library and information science.

In a more recent encyclopedia entry on metadata, Greenberg (2010) discusses the etymology of the term:

The etymology of metadata draws from the classical Greek prefix *meta*, which means “after, behind, or higher”; and [from the] Latin word *data*, plural for *datum*, which means “a piece of information” or “something given.” Metadata is, in essence, a piece of information generated after the object of focus has been produced; and the information given is at a higher level (more abstract) than the object. (3610–11)

This definition adds two more aspects to her earlier definition. First, a second temporal aspect is considered alongside the purposeful orientation toward the future use of metadata: the generation of metadata comes after the production of the object. Second, metadata is specified as existing at a different level of abstraction than the object, a level considered higher or more abstract. What is meant by higher-level abstraction is further specified in Greenberg (2005) where metadata, at one point, is defined as addressing “data attributes that describe, provide context, indicate the quality, or document other object (or data) characteristics” (20). Greenberg provides a number of definitions across her body of work and describes important aspects of metadata that do not clearly point toward one overarching definition.

Studies defining metadata usually also offer typologies of metadata. Burnett, Ng, and Park (1999) propose a simple dichotomy between two types of metadata, which closely maps to their two types of characterization: metadata is either “intrinsic (i.e., those that are related to resource identification and discovery) [or] extrinsic (i.e., those that are related [to] administration and nonbibliographic data)” (1215). The early stages of the development of the Dublin Core metadata scheme followed this division. The initial workshop in Dublin, Ohio, resulted in a set of descriptive metadata elements that adhered to, among others, the principle of intrinsicity (Weibel 1995). The subsequent

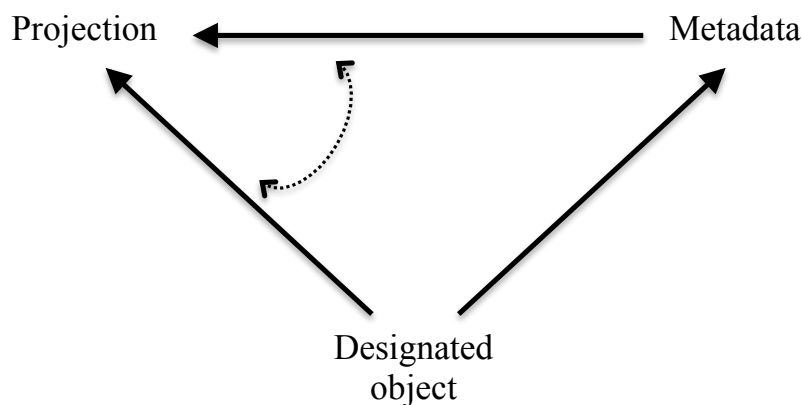
workshop in Warwick, U.K., which sought to create a container architecture for aggregate packages of metadata, also introduced extrinsic metadata elements: terms and conditions, administrative data, content ratings, provenance, linkage or relationship data, and structural data. These types of elements were seen as necessary to enable the scheme to accommodate user communities needing to administer, archive, and control access to data and other information objects (Lagoze 1996). The seven types of metadata identified by the developers of the Dublin Core, usually consolidated into lists of three to five types, makes up the constituents of typologies in later scholarship. For instance, the National Information Standards Organization (NISO 2004) identifies three main types metadata—descriptive, structural, and administrative—and list rights management and preservation as subtypes under administrative metadata. In contrast, Anne Gilliland (2008) settles on five types: administrative, descriptive, preservation, technical, and use.³⁰ Whereas the intrinsic/extrinsic distinction proposed by Burnett, Ng, and Park was based on the idea of attributes inherent to the source object versus its context-dependent uses, later typologies shifted to function as the central principle for distinguishing between types of metadata, as will be explored below.

Describing the foregoing definitions of metadata in a semiotic idiom not only provides new conceptual tools for description, but also introduces a shift in the understanding of what metadata is. Greenberg's (2003) expansive definition of designated objects—"any entity, form, or mode for which contextual data can be

³⁰ See Greenberg (2005, 21), table 1, for a comparison of metadata element typologies.

recorded” (1876, emphasis removed)³¹—provides a useful place to begin a semiotic analysis. Analyzed as a third, what Greenberg points to as the designated object is a sign, the contextual data (metadata) is an interpretant, and a correspondence-preserving projection is the semiotic object (figure 6).³²

Figure 6. Metadata Generation as a Third.



What are the implications of this model? First, it fundamentally changes the understanding of the object of metadata by analytically splitting it between Peircean sign and semiotic object. The “radically diverse and seeming endless” possible objects of metadata to which Greenberg (2003) gestures, “ranging from corporeal and digital information resources, such as a monograph, newspaper, or photograph—to activities, events, persons, places, structures, transactions, relationships, execution directions, and programmatic applications” (1876, emphasis removed), are recognizable as the array of Peircean signs. Greenberg’s candidate for an all-encompassing definition of a designated

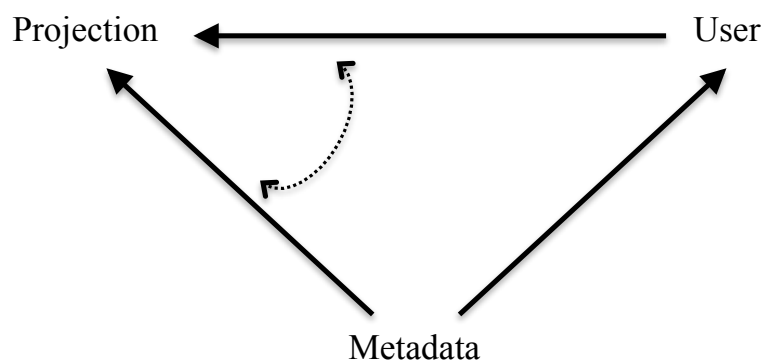
³¹ Note that this definition holds for the process of generating metadata. The importance of point-of-view, or the roles of those who engage with metadata, will be discussed below.

³² To reduce confusion during this section, objects of a third will be called “semiotic objects” to distinguish them from Greenberg’s “designated objects” and “possible objects of metadata,” which occupy different constituents of thirds in the following models.

object, ISO/IEC's (1999) 11179-1 Specification and Standardization of Data Elements, which defines an object as "any part of the conceivable or perceivable world" (6), might hold as an all-encompassing definition of a Peircean sign.³³ The important point for the model of a third, however, is that these definitions describe the sign for which metadata is an interpretant.

The semiotic model of metadata use differs from that of metadata generation. During metadata use, the designated object of metadata generation is no longer part of the picture: the metadata is the sign, the correspondence-preserving projection is the semiotic object, and the metadata user is the interpretant (figure 7). Users of metadata could be several different kinds of interpretants including the mind of an information-seeker, a librarian, an XML parser, or a DNS resolver.

Figure 7. Metadata Use as a Third.

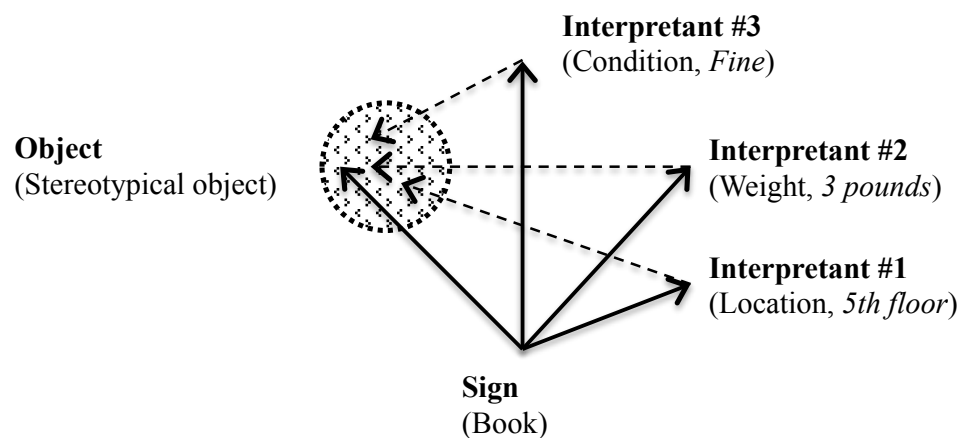


What kinds of correspondence-preserving projections do these different kinds of designated objects and metadata, as signs, create? At least three different types of semiotic objects can be distinguished: (1) stereotypical objects; (2) the text objectified by

³³ Kockelman (2005) includes embodied signs within his list of possible signs, which are only perceivable through patterns of signification and interpretation. Intentional statuses, for instance, would need to be embodied in a perceivable medium and read within a particular semiotic frame for conceivables to become signs.

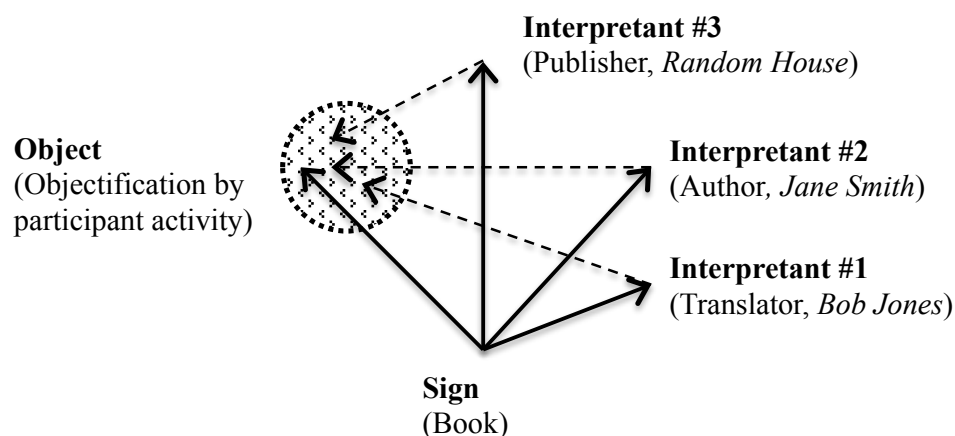
participant activity; (3) the text objectified as intellectual content. Kockelman (2006b) describes stereotypical objects as those “things continuously present to the senses, detachable from context, relative[ly] portable across context, and handy relative to the size and strength of humans” (243) as defined in James Gibson’s (1979) perceptual ecology. They might also be described as the referent of a signifier, as understood by Saussure (1966). A stereotypical example of such a stereotypical object within library science would be a book. The metadata fields of a bibliographic record include a number of the book’s possible qualities as a physical artifact, taken as signs: the (actually existing) book’s weight, dimensions, number of pages, and physical condition, as well as its spatial location are sinisigns. The description of the book’s condition is a replica, a token of a type of condition (e.g., fine, good, fair). The spatial location could be seen as a legisign if its shelf space was determined by a library classification, but the contingent location of the actual book is a sinisign. The book is, to this point, taken as a complex object made from a bundle of possible qualities (figure 8). Notice that other possible qualities, such as the color of the cover, paper weight, font, or width of margins, are not singled out of the bundle for inclusion in the bibliographic record.

Figure 8. Semiotic Object 1: Stereotypical Object.



The same bibliographic record would also include qualities that had nothing to do with the book's material constitution as an artifact. A number of the actors that brought the book into being are included among the metadata fields. Those responsible for the propositional/conceptual content of the book, such as the author, editor, or contributor, as well as those with relatively less responsibility for this content, such as a compiler or translator, are included in the record. A publisher or issuing agency, who bears a different responsibility for the book's conceptual content, and physical and aesthetic qualities, is also included on the record. The participant roles of speakers described by Erving Goffman (1981) help to characterize the attributions of responsibility and justify the inclusion of particular agents in a bibliographic record. According to Goffman's typology of participant roles, an animator produces the physical expression, an author selects the words and phrases, and a principle holds the final responsibility for the ideas expressed. For example, when Barak Obama delivers a speech, he is recognized as the animator and principle, even though his speechwriters acted as authors. In our bibliographic record example, the author of the book is ascribed the role of principle and author, while an editor, compiler, or translator would be seen more as an author and less a principle. Publishers are part principle, part animator. Other animators, such as typesetters, printers, papermakers, and designers are not included in the bibliographic record, or rather they are folded into the corporate personhood of the publisher. Whatever the participant roles of the various actors included in the bibliographic record, their collective presence on the record contributes to the coherence of the object as a complex material/conceptual entity (figure 9).

Figure 9. Semiotic Object 2: Objectification by Participant Activity.



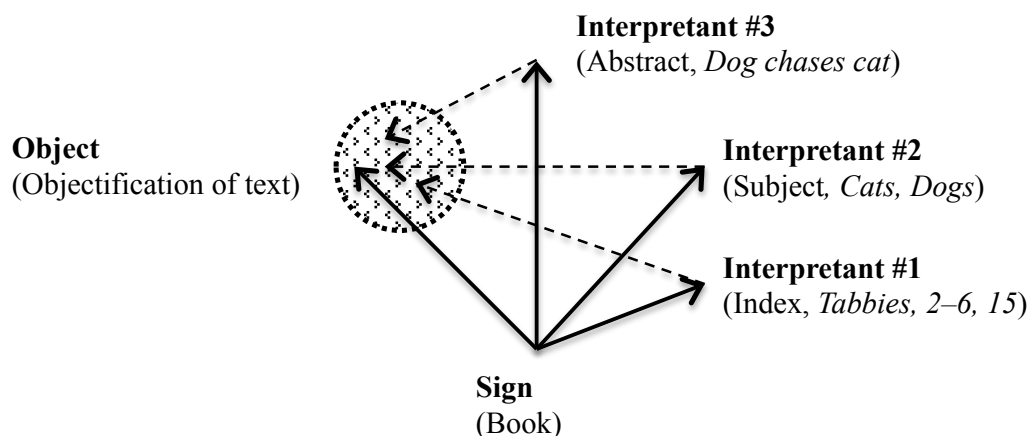
The inclusion of particular actors with specific forms of responsible for bringing about the designated object points to an important aspect of correspondence-preserving projections: they entail a process of objectification.³⁴ Those actors chosen for inclusion contribute not only to the production of the book as a stereotypical object or text-artifact, but also to its objectification as a text, defined by Hanks (1989) as a “configuration of signs that is coherently interpretable by some community of users” (95). The stability or coherence of a text rests on what Hanks labels as textuality, “the fit between the sign form and some larger context” (96). Irrespective of any text-internal cohesion, the inclusion of principles, authors, and animators in a bibliographic record works to cohere and identify a portion of social life, and then to objectify the result (figure 10).³⁵

³⁴ See Keane (2003a) for a discussion of the need to generalize the concept of objectification as a social process beyond Marxist and Heideggerian critiques of reification and inauthenticity.

³⁵ Compare Hanks’s (1989) account of texts and textuality with Richard Smiraglia’s (2001) tripartite distinction between document, text, and work in bibliographic records. While Smiraglia is primarily concerned with defining the nature of works as the products of authorial intention, the distinction he sets up between document and text, where documents are physical containers and texts are “sets of words that constitute writing” (3), is highly problematic. The words “in” a document, as signs, must be perceivable to the senses, presumably as material marks on the page.

Silverstein and Urban (1996) speculate that Paul Ricoeur’s (1981 [1971]) influential essay authorized the conflation of texts and text-artifacts for many interpretive scholars in the

Figure 10. Semiotic Object 3: Objectification of Text.



While Hanks is concerned with the combination of sign complex and cultural context that cohere texts, Silverstein and Urban (1996) are more interested in the process by which parts of discourse are segmented and extracted from ongoing social life and later re-inserted in new contexts: “the process that result in phenomenal textuality—what we have come to call the simultaneous processes of ‘entextualization’ and ‘co(n)textualization’—are the central and ongoing practices within cultural order” (1). The contribution of metadata generators to the cultural order, in the example of bibliographic records, is meta-textual (Hanks 1989, 96): metadata helps to constitute books and other resources as a coherent, entextualized object separated from their originating contexts and therefore able to circulate, as propositional content, to new contexts.

social sciences. They argue, however, that “to confuse the mediating artifact and its mode of production (‘inscription’) for a text and the sociosemiotic process that produce it perpetuates a particular fetishized substitution” (Silverstein and Urban 1996, 2n1).

While the document–text distinction is untenable, works could be approached as embodied signs, with the documents collocated by catalogers standing in for the body of the author. Notice that what Smiraglia (2001) calls documents and works are what Hanks (1989) would call texts and textuality, and what Silverstein and Urban (1996) would call text-artifacts and texts.

Metadata as Function

The array of definitions of metadata proposed by the scholars above differ in a number of ways: in their focus, understanding of the designated object or data, and the basis of the relationship of data to metadata. Of greater interest in most of the literature is not the question of what metadata is, but what metadata is meant to do. Karen Coyle (2010), for instance, states: “It is undoubtedly evident that the creation of good, functional metadata depends greatly on an understanding of the potential uses of the metadata and of the needs that the metadata must be designed to satisfy. . . . [T]he metadata that we find ourselves using everyday is the metadata that we can use to accomplish some task” (6).

As metadata began to proliferate in the closing years of the twentieth-century, library and information science scholars sought to identify and classify metadata elements. The principle on which these typologies came to rely was function. Jane Greenberg (2005) goes so far as to state, after comparing several different typologies, that “the naming of different types of metadata, with labels such as ‘resource discovery’ and ‘use,’ demonstrates that *functionality* is the principle reason for metadata” (22). Some scholars, such as Sherry Vellucci (2000), include functionality within their definitions of metadata: “I use the following definition, which incorporates statements of functionality and environment: Metadata are data that describe the attributes of a resource; characterize its relationships; [and] support its discovery, management, and effective use . . . ” (34).³⁶

³⁶ The last clause of the definition—“and exist in an electronic environment”—was removed. Before stating her definition, Vellucci argues against those librarians who felt that metadata generation in library settings should be considered a type of cataloging. Against this view, she expresses the benefits of adopting terminology used in a wider array of professional communities. It is unclear why Vellucci decided to include the last aspect in her definition.

If metadata's reason for being is to serve particular functions, then the identification and classification of functions becomes the first task for research on metadata, with the definition of element types a derivative secondary task. The metadata types identified at the Warwick conference (Lagoze 1996) should therefore be understood as deriving from the functions the conference participants wanted the Dublin Core to support in addition to resource discovery (table 3).

Table 3. Dublin Core Metadata Functions and Types.

Metadata Functions	Types of Metadata
Resource discovery / information retrieval	Descriptive / information
Resource management	Administrative
Resource use	Terms and conditions
Resource use by appropriate audience	Content ratings
Resource authentication and other provenance-related activities	Provenance
Resource contextualization among related resources	Linking / relationship
Identifying resource hardware and software needs	Structural

Sources: Lagoze (1996). Table adapted from Greenberg (2005, 21) table 1.

The report of the Working Group on the Use of Metadata Schemas of the International Federation of Library Associations and Institutions (IFLA 2005) does not explicitly point to function as the *raison d'être* of metadata, but their typology can also be read as pairing types with functions.³⁷ Written to provide guidance to libraries choosing and implementing metadata schemes, the report includes a number of metadata functions and types falling outside the purview of the Dublin Core (table 4).

³⁷ The chairman of the group, Lynne Howarth (2005), suggests the centrality of functions in the report's typology in an article published a few months after the report: "Metadata schemas have been developed to meet the needs of particular fields or domains . . . and to support a variety of functions" (41).

Table 4. IFLA Metadata Functions and Types.

Metadata Functions	Types of Metadata
Resource / metadata management	Administrative
Resource discovery / information retrieval	Descriptive
Resource analysis and enhanced use	Analytical
Resource use restrictions	Rights management
Tracking resource media conversions	Technical
Resource preservation	Preservation
Resource versioning and compiling	Structural
Tracking of resource use	Use
Locally-determined function not falling under other metadata functions	Other

Sources: IFLA (2005, 7–9); Howarth (2005, 41).

Comparing these two typologies confirms Greenberg’s (2005) observation of the overlaps and differing nomenclature that hinders the development of a universal list of types. It also brings to the fore a number of common characteristics or behaviors of metadata. For instance, both typologies seek to support the collection of information that characterizes not just the designated object, but also the metadata itself. Enabling metadata to document itself by including elements for its creators, or dates of creation and modification, is seen as an essential function. The need for “administrative metadata,” as it is labeled by both typologies, what Greenberg (2003, 1881) identifies as “meta-metadata,” highlights a characteristic of metadata noted by the participants of the Warwick conference: “What may appear to be metadata in one context, may look very much like data in another. . . . This recursive relationship of data to metadata may nest to an arbitrarily deep level” (Lagoze 1996, n.p.). The potential of nested levels, data to metadata to meta-metadata and so on, and the arbitrariness of the levels from one instance to the next, hints at the possible opening out of metadata beyond the immediate horizon of a particular scheme. It also gestures at the tendency of metadata to beget metadata, what speculatively could be understood as a consequence of metadata’s

constitutive insufficiency in fulfilling its documentary function without supplement. Here, meta-metadata should be understood by its functional principle, the need for documentation of the object's surrogate record as a resource in itself.

Matthew Mayernik (2011), in his ethnographic study of data and metadata practices among environmental scientists defines metadata as “documentation, descriptions, and annotations created and used to manage, discover, access, use, share, and preserve informational resources” (28, emphasis removed). Like Greenberg, he focuses on the functions metadata are designed to perform, although for Mayernik these functions are more tightly bound to specific situations. He is interested in the social and institutional contexts in which metadata is created to pursue specific work practices or tasks within the sciences. Comparing ideas about metadata from different settings, he argues that the “principle-based development of metadata schemas and systems” (38) of information institutions is not shared by metadata initiatives in scientific research communities: “Many scientific projects may not require, or benefit from, the kinds of metadata standardization that information institutions have developed for bibliographic and archival description. In these cases, customization and looseness of metadata descriptions are to be expected” (38–39).³⁸

Mayernik's ethnographic focus on the metadata practices of particular groups of scientists required a broader understanding of the functions metadata supports. He points

³⁸ Jeffrey Beall (2005), arguing from the principle-based view of metadata as an academic catalog librarian, warns against what he calls the “Tower of Babel” of metadata schemes, and advocates for the wider adoption of MARC as a single, comprehensive metadata standard. This type of critique would see the ad-hoc metadata practices of many scientists as betraying what Beall understands as the primary function of metadata: to support the possibility of discovery beyond any particular research community.

to Coyle's (2010) discussion of what she identifies as the three general attributes of metadata: it is constructed, constructive, and actionable:

- Constructed: Metadata is not found in nature. It is entirely an invention; it is an artificiality.
- Constructive: Metadata is constructed for some purpose, some activity, to solve some problem. . . .
- Actionable: The point of metadata is to be useful in some way. This means that it is important that one can act on the metadata in a way that satisfies some needs. (6, emphasis removed)

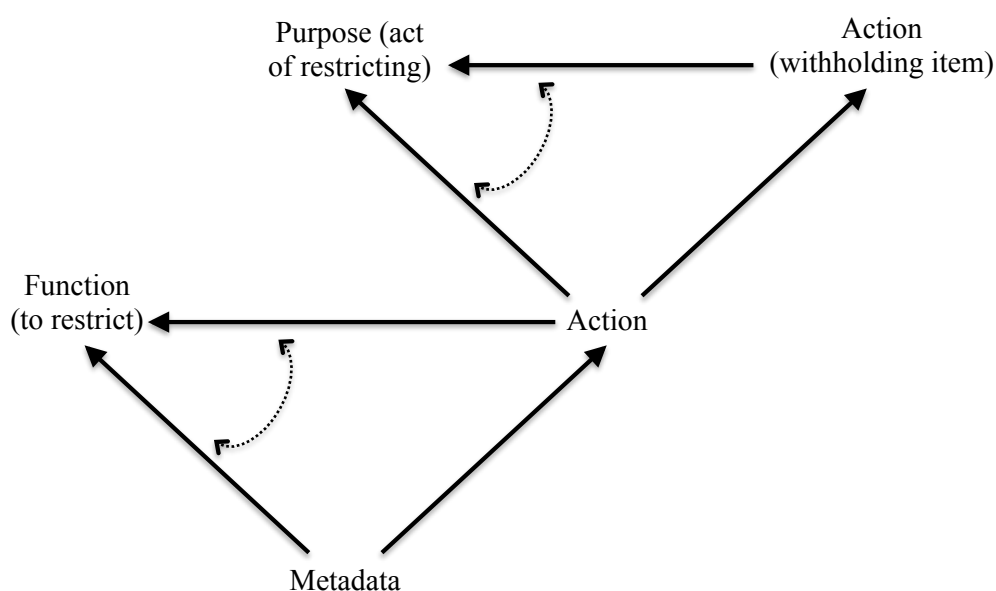
Although presented as a “functional definition [that] gives us a viable roadmap for our own studies of metadata utility and quality” (6), Coyle's list is less a definition of metadata than the identification of the common attributes of any instrument. If an instrument can be defined as an artificed entity whose object is a function (Kockelman 2006b), then by extension, metadata might be defined as an artifact whose object is to fulfill a resource-related function.

Coyle's instrumental description of metadata can be analyzed as a third, or rather, a chain of thirds. Kockelman (2006b) defines an instrument as a “semiotic process whose sign is an artificed entity, whose object is a function, and whose interpretant is an affordance that contextualizes it, another instrument that incorporates it, an action that wields it, or an utterance that represents it” (37). His analysis is aimed at carefully distinguishing the different constituent types of semiosis that work together to constitute a wielded instrument. Importantly, Kockelman understands the process of wielding instruments as one of a number of general, non-propositional semiotic processes, including heeding affordances, undertaking actions, performing roles, and filling identities.

What might this semiotic model of metadata as an instrument look like? Taking an example of a metadata function associated with resource use, say use-restrictions to

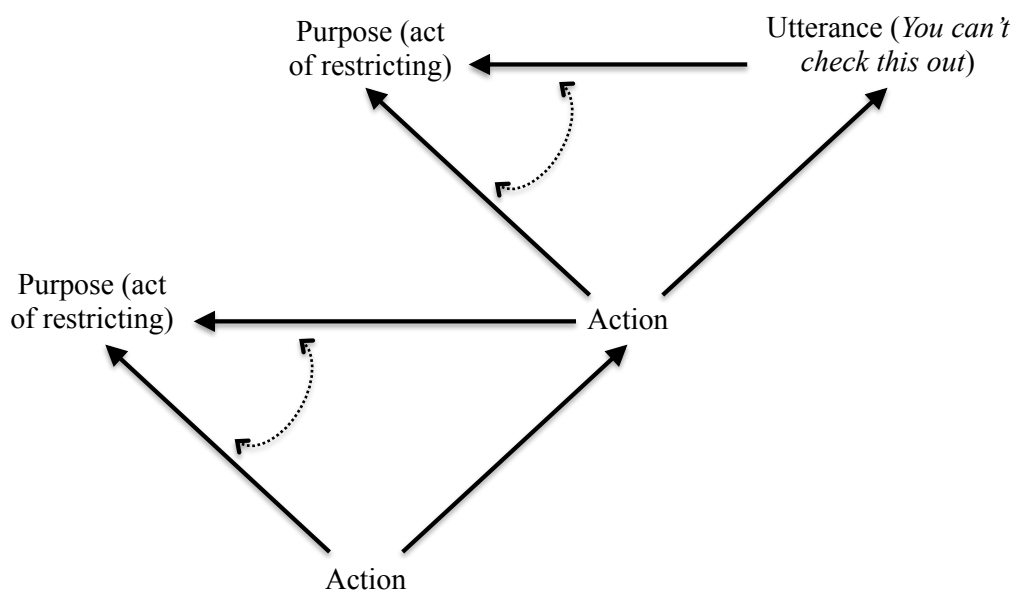
block certain age groups from checking out a resource from the library, a semiotic model would consist of two or more thirds linked in a chain. In the first third, the metadata field (sign) stands for the function of restriction (object), which is realized by an action (interpretant). The function of the use-restrictions can only be achieved through the prompting of an action, a second type of non-propositional semiotic process, whose object is a purpose. Kockelman (2006b) calls this second linked semiotic process a realizing embedded interpretant: “For any two semiotic processes, A and B, A will be said to *realize* B (and hence be interpreted by it) if B is an objectification of the object of A” (30). In our example, the action of withholding (A) realizes the instrumental function of the resource-use metadata (B). The interpretant of the first third is the sign in the second: the controlled behavior (sign) stands for the purpose of restriction (object), which is realized through a secondary controlled behavior, withholding the item (interpretant) (figure 11).

Figure 11. Metadata as Instrument Realized by an Action (Resource-use example).



In our library scenario, the action of withholding might be paired with other actions such as an utterance: “You can’t check this out.” If verbally informing a patron is considered integral to the action of restricting the use of a resource, then this action is the realizing interpretant of the nonverbal action of withholding (figure 12).

Figure 12. Metadata as Instrument Realized by an Action Realized by an Utterance (Resource-use example).



In this example, the interpretant of the controlled behavior is a performative utterance (Austin 2003 [1955]). In Kockelman’s (2005) terms, for the action to achieve its purpose the utterance must be appropriate in context and effective on context. The felicity of this particular utterance depends on already-established norms of librarian–patron interaction: the authority of a librarian to determine who can check out what resource. For this particular interaction to be appropriate in context it must be recognized as a token of a type of interaction, and to be effective on context it must set up subsequent interactions as tokens of this type. The links between semiotic processes, norms of interaction (tokens),

and particular contexts (types) that enable metadata to fulfill its functions must be traced to understand how metadata works as an instrument.

Conclusion

This thesis has reviewed the literatures of library and information science and science studies to identify the constituents and relations of metadata as a complex object and process. It has also taken the first steps of articulating a unified analytic framework for approaching metadata as signification that establishes common ground, adapting semiotic anthropology for this purpose. Several examples of the kind of analysis made possible by this framework were used as demonstrations, although they stop far short of their full potential. In particular, the examples analyze relatively more pragmatic and less semantic aspects of metadata. The analysis of the objectification of text, for instance, highlights the part metadata plays in stabilizing the identity of a text as an object. It does not tackle the thornier issues of representation, what Campbell (2005) centrally confronts with his analysis of metonymy. Kockelman (2013a) has sought to address the relationship of representational and non-representational semiotic processes and their interactions, and this is the direction in which a semiotic-functional analysis of metadata must proceed in the future.

If the initial gesture of making an analogical connection between disciplines was bold, the difficulties of transforming analogy into a working analysis encourages a timid conclusion. This thesis hopefully validates the potential of a semiotic anthropological approach to central issues in information organization and data mobility. It might also serve to direct anthropological attention to a phenomenon of increasing importance in the contemporary world.

References

- Agha, Asif. 2007a. *Language and Social Relations*. New York: Cambridge University Press.
- . 2007b. “The Object Called ‘Language’ and the Subject of Linguistics.” *Journal of English Linguistics* 35, no. 3: 217–35.
- Akrich, Madeleine, and Bruno Latour. 1992. “A Summary of a Convenient Vocabulary for the Semiotics of Human and Nonhuman Assemblies.” In *Shaping Technology/Building Society: Studies in Sociotechnical Change*, edited by Wiebe E. Bijker and John Law, 259–264. Cambridge, Mass.: MIT Press.
- Almklov, Petter G. 2008. “Standardized Data and Singular Situations.” *Social Studies of Science* 38, no. 6: 873–97.
- American Council of Learned Societies (ACLS). 2006. *Our Cultural Commonwealth: The Report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences*. Available from ACLS website: http://www.acls.org/uploadedFiles/Publications/Programs/Our_Cultural_Commonwealth.pdf.
- Atkins, Daniel E., Kelvin K. Droegemeier, Stuart I. Feldman, Hector García-Molina, Michael L. Klein, David G. Messerschmitt, Paul Messina, Jeremiah P. Ostriker, Margaret H. Wright. 2003. *Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon*

- Advisory Panel on Cyberinfrastructure*. National Science Foundation (NSF).
Available from NSF website: <http://www.nsf.gov/od/oci/reports/atkins.pdf>.
- Austin, J. L. 2003 [1955]. *How to Do Things with Words*, 2nd ed. Cambridge, Mass.: Harvard University Press.
- Baker, Karen, David Ribes, Florence Millerand, Geoffrey C. Bowker. 2005.
“Interoperability Strategies for Scientific Cyberinfrastructure: Research and Practice.” *Proceedings of the American Society for Information Science and Technology* 42, no. 1.
- Bauman, Richard, and Charles L. Briggs. 1990. “Poetics and Performance as Critical Perspectives on Language and Social Life.” *Annual Review of Anthropology* 19: 59–88.
- Bergman, Manfred Max, and Thomas S. Eberle, eds. 2005. “Qualitative Inquiry: Research, Archiving, and Reuse.” Special issue, *Forum: Qualitative Social Research* 6, no. 2.
- Bernstein, Richard J. 1983. *Beyond Objectivism and Relativism: Science, Hermeneutics, and Praxis*. Philadelphia: University of Pennsylvania Press.
- Berry, David M. 2011. “The Computational Turn: Thinking about the Digital Humanities.” *Culture Machine* 12: 1–22.
- Blair, D. C. 1990. *Language and Representation in Information Retrieval*. New York: Elsevier.
- Blok, Anders, and Torben Elgaard Jensen. 2011. *Bruno Latour: Hybrid Thoughts in a Hybrid World*. New York: Routledge.

- Borgman, Christine L. 2012. "The Conundrum of Sharing Research Data." *Journal of the American Society for Information Science and Technology* 63, no. 6: 1059–78.
- . 2007. *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. Cambridge, Mass.: MIT Press.
- Borgman, Christine L., Jillian C. Wallis, and Noel Enyedy. 2007. "Little Science Confronts the Data Deluge: Habitat Ecology, Embedded Sensor Networks, and Digital Libraries." *International Journal of Digital Libraries* 7, nos. 1–2: 17–30.
- Bowker, Geoffrey C. 2007. *Memory Practices in the Sciences*. Cambridge, Mass.: MIT Press.
- . 2000. "Biodiversity Datadiversity." *Social Studies of Science* 30, no. 5: 643–83.
- Bowker, Geoffrey C., Karen Baker, Florence Millerand, and David Ribes. 2010. "Towards Information Infrastructure Studies: Ways of Knowing in a Networked Environment." In *International Handbook of Internet Research*, edited by Jeremy Hunsinger, Lisbeth Klastrup, and Matthew Allen, 97–117. New York: Springer.
- Bowker, Geoffrey C., and Susan Leigh Star. 1999. *Sorting Things Out: Classification and Its Consequences*. Cambridge, Mass.: MIT Press.
- boyd, danah, and Kate Crawford. 2012. "Critical Questions for Big Data." *Information, Communication & Society* 15, no. 5: 662–79.
- Brier, Søren. 2008. *Cybersemiotics: Why Information Is Not Enough*. Toronto: University of Toronto Press.
- . 2006. "The Foundation of LIS in Information Science and Semiotics." *Libreas: Library Ideas* 1, no. 2.

- Brubaker, Rogers. 1993. "Social Theory as Habitus." In *Bourdieu: Critical Perspectives*, edited by Craig Calhoun, Edward LiPuma, and Moishe Postone, 212–34. Chicago: University of Chicago Press.
- Buckland, Michael 1997. "What Is a 'Document'?" *Journal of the American Society for Information Science and Technology* 48, no. 9: 804–809.
- Burawoy, Michael. 1998. "The Extended Case Method." *Sociological Theory* 16(4), 4–34.
- Burnett, Kathleen, Kwong Bor Ng, and Soyeon Park. 1999. "A Comparison of the Two Traditions of Metadata Development." *Journal of the American Society for Information Science* 50, no. 13: 1209–17.
- Campbell, D. Grant. 2005. "Metadata, Metaphor, and Metonymy." *Cataloging & Classification Quarterly* 40, no. 3–4: 57–73.
- Campbell, Eric G., Brian R. Clarridge, Manjusha Gokhale, Lauren Birenbaum, Stephen Hilgartner, Neil A. Holtzman, David Blumenthal. 2002. "Data Withholding in Academic Genetics: Evidence from a National Survey." *Journal of the American Medical Association* 287, no. 4: 473–80.
- Carlson, Samuelle, and Ben Anderson. 2007. "What *Are* Data? The Many Kinds of Data and Their Implications for Data Re-Use." *Journal of Computer-Mediated Communication* 12: 635–51.
- Case, Donald O. 2007. "Models of Information Behavior." In *Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior*, 2nd ed., 119–40. Bingley, U.K.: Emerald.

- Caton, Steven C. 1987. "The Contributions of Roman Jakobson." *Annual Review of Anthropology* 16: 223–60.
- Chakrabarty, Dipesh. 2000. *Provincializing Europe: Postcolonial Thought and Historical Difference*. Princeton, N.J.: Princeton University Press.
- Coopmans, Cateljine. 2006. "Making Mammograms Mobile: Suggestions for a Sociology of Data Mobility." *Information, Communication & Society* 9, no. 1: 1–19.
- Coyle, Karen. 2010. "Library Data in a Modern Context." *Library Technology Reports* 46, no. 1: 5–13.
- Dreyfus, Herbert, and Paul Rabinow. 1983. *Michel Foucault: Beyond Structuralism and Hermeneutics*. 2nd. ed. Chicago: University of Chicago Press.
- Duranti, Alessandro. 2003. "Language as Culture in U.S. Anthropology." *Current Anthropology* 44, no. 3: 323–47.
- Eco, Umberto. 1986. *Semiotics and the Philosophy of Language*. Bloomington: Indiana University Press.
- . 1979. *Theory of Semiotics*. Bloomington: Indiana University Press.
- Edwards, Paul N. 2010. *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*. Cambridge, Mass.: MIT Press.
- Edwards, Paul N., Geoffrey C. Bowker, Stephen J. Jackson, Robin Williams. 2009. "Introduction: An Agenda for Infrastructure Studies." *Journal of the Association for Information Systems* 10, no. 5: 364–74.
- Edwards, Paul N., Steven J. Jackson, Geoffrey C. Bowker, and Cory P. Knobel. 2007. "Understanding Infrastructure: Dynamics, Tensions, and Design." Report of a Workshop on "History and Theory of Infrastructure: Lessons for New Scientific

- Cyberinfrastructure.” National Science Foundation. Available at <http://hdl.handle.net/2027.42/49353>.
- Edwards, Paul N., Matthew S. Mayernik, Archer L. Batcheller, Geoffrey C. Bowker, and Christine L. Borgman. 2011. “Science Friction: Data, Metadata, and Collaboration.” *Social Studies of Science* 41, no. 5: 667–90.
- Engelke, Matthew. 2007. *The Problem of Presence: Beyond Scripture in an African Church*. Berkeley: University of California Press.
- Espland, Wendy Nelson, and Mitchell L. Stevens. 1998. “Commensuration as a Social Process.” *Annual Review of Sociology* 24: 313–43.
- Fabian, Johannes. 2008. *Ethnography as Commentary: Writing from the Virtual Archive*. Durham, N.C.: Duke University Press.
- Faniel, Ixchel M., and Trond E. Jacobsen. 2010. “Reusing Scientific Data: How Earthquake Engineering Researches Assess the Reusability of Colleagues’ Data.” *Computer Supported Cooperative Work* 19, nos. 3–4: 355–75.
- Frohmann, Bernd. 2004. *Deflating Information: From Science Studies to Documentation*. Toronto: University of Toronto Press.
- Frye, Northrop. 1977. *Anatomy of Criticism: Four Essays*. Princeton, N.J.: Princeton University Press.
- Fujimura, Joan H. 1992. “Crafting Science: Standardized Packages, Boundary Objects, and ‘Translation.’” In *Science as Practice and Culture*, edited by Andrew Pickering, 168–211. Chicago: University of Chicago Press.
- Gadamer, Hans-Georg. 1985 [1965]. *Truth and Method*. Translated by Garret Barden and John Cumming. New York: Crossroad.

- Gaonkar, Dilip P., and Elizabeth A. Povinelli. 2003. "Technologies of Public Forms: Circulation, Transfiguration, Recognition." *Public Culture* 15, no. 3: 385–98.
- Gerson, Elihu, and Susan Leigh Star. 1986. "Analyzing Due Process in the Workplace." *ACM Transactions on Office Information Systems* 4, no. 3: 257–70.
- Gibson, James J. 1979. *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.
- Gilliland, Anne J. 2008. "Setting the Stage." In *Introduction to Metadata 3.0*, edited by Murtha Baca. Los Angeles: Gerry Publications. Available at Getty website: http://www.getty.edu/research/publications/electronic_publications/intrometadata/index.html.
- Goffman, Erving 1981. "Footing." In *Forms of Talk*, 124–59. Philadelphia: University of Pennsylvania Press.
- Gray, Jim. 2009. "Jim Gray on eScience: A Transformed Scientific Method." In *The Fourth Paradigm: Data-Intensive Scientific Discovery*, edited by Tony Hey, Stewart Tansley, and Kristin Tolle, xvi–xxxi. Redmond, Wash.: Microsoft Research.
- Greenberg, Jane. 2010. "Metadata and Digital Information." In *Encyclopedia of Library and Information Science*, 3rd ed., 3610–23. New York: Taylor & Francis.
- . 2005. "Understanding Metadata and Metadata Schemes." *Cataloging & Classification Quarterly* 40, nos. 3–4: 17–36.
- . 2003. "Metadata and the World Wide Web." In *Encyclopedia of Library and Information Science*, edited by M. Drake, 1876–88. New York: Marcel Dekker.

- Hankins, Joseph D. 2013. "Semiotics of Organizations." In *A Companion to Organization Anthropology*, edited by D. Douglas Caulkins and Ann T. Jordan, 204–18. New York: Blackwell.
- Hanks, William F. 1996. *Language and Communicative Practice*. Boulder, Colo.: Westview.
- . 1989. "Text and Textuality." *Annual Review of Anthropology* 18: 95–127.
- Heery, Rachel. 1996. "Review of Metadata Formats." *Program: Electronic Library and Information Systems* 30, no. 4: 345–75.
- Hess, David J. 1998. "If You're Thinking of Living in STS: A Guide for the Perplexed." In *Cyborgs and Citadels: Anthropological Interventions in Emerging Sciences and Technologies*, edited by Gary Lee Downey and Joseph Dumit, 143–64. Santa Fe, N.M.: SAR Press.
- Høstaker, Roar. 2005. "Latour—Semiotics and Science Studies." *Science Studies* 18, no. 2: 5–25.
- Howarth, Lynne C. 2005. "Metadata and Bibliographic Control: Soul-Mates or Two Solitudes?" *Cataloging & Classification Quarterly* 40, nos. 3–4: 37–56.
- Hughes, Thomas P. 1989. "Evolution of Large Technological Systems." In *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*, edited by Wiebe E. Bijker, Thomas P. Hughes, Trevor Pinch, 51–82. Cambridge, Mass.: MIT Press.
- Hull, Matthew. 2012. "Documents and Bureaucracy." *Annual Review of Anthropology* 41: 251–67.

International Federation of Library Associations and Institutions (IFLA). 2005.

“Guidance on the Nature, Implementation, and Evaluation of Metadata Schemas in Libraries.” Final Report of the IFLA Cataloguing Section Working Group on the Use of Metadata Schemas. Available from IFLA website:
http://www.ifla.org/files/assets/cataloguing/pubs/metadata_schemas-20050731.pdf.

International Organization for Standardization / International Electrotechnical Commission (ISO/IEC). 1999. Information Technology—Specification and Standardization of Data Elements, Part 1. 11179-1. International standard. Geneva: ISO Copyright Office.

Irvine, Judith. 1989. “When Talk Isn’t Cheap: Language and Political Economy.” *American Ethnologist* 16, no. 2: 248–267.

Jakobson, Roman. 1990. *On Language*, edited by L. R. Waugh and M. Monville-Burston. Cambridge, Mass.: Harvard University Press.

———. 1971a [1966]. “Relationship Between Russian Stem Suffixes and Verbal Aspects. In *Selected Writings*, vol. 2, *Word and Language*, 198–202. Hague: Mouton.

———. 1971b [1957]. “Shifters and Verbal Categories, and the Russian Verb.” In *Selected Writings*, vol. 2, *Word and Language*, 130–47. Hague: Mouton.

———. 1960. “Closing Statement: Linguistics and Poetics.” In *Style in Language*, edited by Thomas A. Sebeok, 350–77. New York: John Wiley & Sons.

Jirotko, Marina, Rob Procter, Mark Hartswood, Roger Slack, Andrew Simpson, Catelijne Coopmans, Chris Hinds, and Alex Voss. 2005. “Collaboration and Trust in

- Healthcare Innovation: The eDiaMoND Case Study.” *Computer Supported Cooperative Work* 14, no. 4: 369–98.
- Keane, Webb. 2003a. “Self-Interpretation, Agency, and the Objects of Anthropology: Reflections on a Genealogy.” *Comparative Studies in Society and History* 45, no. 2: 222–48.
- . 2003b. “Semiotics and the Social Analysis of Material Things.” *Language & Communication* 23, nos. 3–4: 409–25.
- . 1997. *Signs of Recognition: Power and Hazards of Representation in an Indonesian Society*. Berkeley: University of California Press.
- Knorr-Cetina, Karin. 1981. *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science*. New York: Pergamon.
- Kockelman, Paul. 2013a. *Agent, Person, Subject, Self: A Theory of Ontology, Interaction, and Infrastructure*. New York: Oxford University Press.
- . 2013b. “Information is the Enclosure of Meaning: Cybernetics, Semiotics, and Alternative Theories of Information.” *Language & Communication* 33, no. 2: 115–27.
- . 2006a. “A Semiotic Ontology of the Commodity.” *Journal of Linguistic Anthropology* 16, no. 1: 76–102.
- . 2006b. “Residence in the World: Affordances, Instruments, Actions, Roles, and Identities.” *Semiotica* 162, no. 1: 19–71.
- . 2005. “The Semiotic Stance.” *Semiotica* 157, no. 1: 233–304.

- Kockelman, Paul, and Anya Bernstein. 2013. "Semiotic Technologies, Temporal Reckoning, and the Portability of Meaning. Or: Modern Modes of Temporality—Just How Abstract are They?" *Anthropological Theory* 12, no. 3: 320–48.
- Kuhn, Thomas. 1996 [1962]. *The Structure of Scientific Revolutions*. 3rd ed. Chicago: University of Chicago Press.
- Lagoze, Carl. 1996. "The Warwick Framework: A Container Architecture for Diverse Sets of Metadata." *D-Lib Magazine* 2, no. 8. Retrieved March 15, 2013, <http://www.dlib.org/dlib/july96/lagoze/07lagoze.html>.
- Law, John. 2009. "Actor Network Theory and Material Semiotics," in *The New Blackwell Companion to Social Theory*, edited by Bryan S. Turner, 141–58. New York: Blackwell.
- Latour, Bruno. 1988. *The Pasteurization of France*. Translated by Alan Sheridan and John Law. Cambridge, Mass.: Harvard University Press.
- . 1987. *Science in Action*. Cambridge, Mass.: Harvard University Press.
- . 1986. "Visualization and Cognition: Thinking with Eyes and Hands." In *Knowledge and Society: Studies in the Sociology of Culture Past and Present* 6: 1–40. Greenwich, Conn.: JAI Press.
- Latour, Bruno, and Steve Woolgar. 1979. *Laboratory Life: The Social Construction of Scientific Facts*. Beverly Hills: Sage Publications.
- Lee, Benjamin. 1997. *Taking Heads: Language, Metalanguage, and the Semiotics of Subjectivity*. Durham, N.C.: Duke University Press.
- Lenoir, Timothy. 1994. "Was the Last Turn the Right Turn? The Semiotic Turn and A. J. Greimas." *Configurations* 2, no. 1: 119–36.

- Lenoir, Timothy, ed. 1997. *Inscribing Science: The Cultural Production of Scientific Disciplines*. Stanford, Calif.: Stanford University Press.
- Locke, Karen. 2010. "Abduction." In *Encyclopedia of Case Study Research*, edited by Albert J. Mills, Gabrielle Durepos, and Elden Wiebe, 2–4. Thousand Oaks, Calif.: Sage.
- Lucy, John A. 1993. "Reflexive Language and the Human Disciplines." In *Reflexive Language: Reported Speech and Metapragmatics*, edited by John A. Lucy, 9–32. New York: Cambridge University Press.
- Lynch, Michael. 1985. *Art and Artifact in Laboratory Science: A Study of Shop Work and Shop Talk in a Research Laboratory*. London: Routledge & Kegan Paul.
- Mai, Jens-Erik. 2001. "Semiotics and Indexing: An Analysis of the Subject Indexing Process." *Journal of Documentation* 57, no. 5: 591–622.
- Manning, Paul. 2012. *The Semiotics of Drink and Drinking*. New York: Continuum.
- Manning, Paul, ed. 2003. "Words and Beyond: Linguistic and Semiotic Studies of Sociocultural Order." Special issue, *Language & Communication* 23, nos. 3–4.
- Manovich, Lev. 2012. "Trending: The Promises and Challenges of Big Social Data." In *Debates in the Digital Humanities*, edited by Matthew K. Gold, 460–75. Minneapolis: University of Minnesota Press.
- Mayernik, Matthew S. 2011. *Metadata Realities for Cyberinfrastructure: Data Authors as Metadata Creators*. PhD dissertation, University of California, Los Angeles.
- Mertz, Elizabeth. 2007. "Semiotic Anthropology." *Annual Review of Anthropology* 36: 337–53.

- Mertz, Elizabeth, and Richard J. Parmentier, eds. 1985. *Semiotic Mediation: Sociocultural and Psychological Perspectives*. Orlando: Academic Press.
- Millerand, Florence, and Geoffrey C. Bowker. 2009. "Metadata Standards: Trajectories and Enactment in the Life of an Ontology." In *Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life*, edited by Susan Leigh Star and Martha Lampland, 149–76. Ithaca, N.Y.: Cornell University Press.
- National Information Standards Organization (NISO). 2004. *Understanding Metadata*. Bethesda, Md.: NISO Press. Available from NISO website: www.niso.org.
- Parmentier, Richard. 1994. *Signs in Society: Studies in Semiotic Anthropology*. Bloomington: Indiana University Press.
- Peirce, Charles Sanders. 1955a. "The Principles of Phenomenology." In *Philosophical Writings of Peirce*, edited by Justus Buchler, 74–97. Mineola, N.Y.: Dover.
- . 1955b. "Logic as Semiotic: The Theory of Signs." In *Philosophical Writings of Peirce*, edited by Justus Buchler, 98–119. Mineola, N.Y.: Dover.
- Peters, John Durham. 1999. *Speaking into the Air: A History of the Idea of Communication*. Chicago: University of Chicago Press.
- Raber, Douglas, and John M. Budd. 2003. "Information as Sign: Semiotics and Information Science." *Journal of Documentation* 59, no. 5: 507–22.
- Rabinow, Paul, and William M. Sullivan, eds. 1987. *Interpretive Social Science: A Second Look*. Berkeley: University of California Press.
- Ragin, Charles C. 1994. *Constructing Social Research*. Thousand Oaks, Calif.: Pine Forge Press.

- Ribes, David, and Thomas A. Finholt. 2009. "The Long Now of Technology Infrastructure: Articulating Tensions in Development." *Journal of the Association of Information Systems* 10, no. 5: 375–98.
- Ribes, David, and Charlotte P. Lee. 2010. "Sociotechnical Studies of Cyberinfrastructure and e-Research: Current Themes and Future Trajectories." *Computer Supported Cooperative Work* 19, nos. 3–4: 231–44.
- Ricoeur, Paul. 1981 [1971]. "The Model of the Text: Meaningful Action Considered as a Text. In *Hermeneutics and the Human Sciences: Essay on Language, Action, and Interpretation*, edited and translated by John. B. Thompson, 197–221. New York: Cambridge University Press.
- Robin, Richard. 1967. *Annotated Catalogue of the Papers of Charles S. Peirce*. Amherst: University of Massachusetts Press.
- Rosenzweig, Roy. 2003. "Scarcity or Abundance? Preserving the Past in the Digital Era." *American Historical Review* 108, no. 3: 735–62.
- Šauperyl, Alenka. 1999. *Subject Determination During the Cataloging Process*. PhD dissertation, University of North Carolina, Chapel Hill.
- de Saussure, Ferdinand. 1966 [1915]. *Course in General Linguistics*, translated by Wade Baskin, edited by Charles Bally and Albert Sechehaye. New York: McGraw-Hill.
- Sebeok, Thomas A. 2001. *Signs: An Introduction to Semiotics*. 2nd ed. Toronto: University of Toronto Press.
- Silverstein, Michael. 1993. "Metapragmatic Discourse and Metapragmatic Function." In *Reflexive Language: Reported Speech and Metapragmatics*, edited by John A. Lucy, 33–58. New York: Cambridge University Press.

- . 1976. “Shifters, Linguistic Categories and Cultural Description.” In *Meaning in Anthropology*, edited by Keith H. Basso and Henry A. Selby, 11–55.
Albuquerque: University of New Mexico Press.
- Silverstein, Michael, and Greg Urban, eds. 1996. *Natural Histories of Discourse*.
Chicago: University of Chicago Press.
- Smiraglia, Richard P. 2001. *The Nature of “A Work”: Implications for the Organization of Knowledge*. Lanham, Md.: Scarecrow Press.
- Star, Susan Leigh, and Karen Ruhleder. 1996. “Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces.” *Information Systems Research* 7, no. 1: 111–34.
- Suchman, Lucy. 1994. “Working Relations of Technology Production and Use.” *Computer Supported Cooperative Work* 2, nos. 1–2: 21–39.
- Tavory, Iddo, and Stefan Timmermans. 2009. “Two Cases of Ethnography: Grounded Theory and the Extended Case Method.” *Ethnography* 10, no. 3: 243–63.
- Taylor, Charles. 1985. *Philosophy and the Human Science: Philosophical Papers 2*. New York: Cambridge University Press.
- Tellefsen, Torkild L. “Semiotic Knowledge Organization: Theory and Method Development.” *Semiotica* 142, no. 1: 71–90.
- Timmermans, Stefan, and Iddo Tavory. 2012. “Theory Construction in Qualitative Research: From Grounded Theory to Abductive Analysis.” *Sociological Theory* 30, no. 3: 167–86.
- Turner, Stephen. 2008. “The Social Study of Science before Kuhn.” In *The Handbook of Science and Technology Studies*, 3rd ed., edited by Edward J. Hackett, Olga

- Amsterdamska, Michael E. Lynch, & Judy Wajcman, 33–62. Cambridge, Mass.: MIT Press.
- Ure, Jenny R., Rob Procter, Yuwei Lin, Mark Hartswood, Stuart Anderson, Sharon Lloyd, Joanna Wardlow, Kate Ho, and Horacio Gonzalez-Velez. 2009. “The Development of Data Infrastructures for eHealth: A Socio-Technical Perspective.” *Journal of the Association for Information Systems* 10, no. 5: 415–19.
- van House, Nancy A. 2004. “Science and Technology Studies and Information Studies.” *Annual Review of Information Science and Technology* 38: 3–86.
- Vellucci, Sherry L. 2000. “Metadata and Authority Control.” *Library Resources & Technical Services* 44, no. 1: 33–43.
- Vertesi, Janet, and Paul Dourish. 2011. “The Value of Data: Considering the Context of Production in Data Economies.” In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work (CSCW '11)*, 533–42. New York: ACM.
- Waldstein, Maxim. 2008. “The Mangle of Practice or the Empire of Signs? Toward a Dialogue Between Science Studies and Soviet Semiotics,” in *The Mangle in Practice: Science, Society, Becoming*, edited by Andrew Pickering and Keith Guzik, 221–42. Durham, N.C.: Duke University Press.
- Warner, Julian. 1990. “Semiotics, Information Science and Computers.” *Journal of Documentation* 46, no. 1: 16–32.
- Watts, Christopher M. 2008. “On Mediation and Material Agency in the Peircean Semeiotic.” In *Material Agency: Towards a Non-Anthropocentric Approach*, edited by Carl Knappett and Lambros Malafouris, 187–207. New York: Springer.

- Weibel, Stuart. 1995. "Metadata: The Foundations of Resource Description." *D-Lib Magazine* 1, no. 1. Retrieved August 29, 2012, <http://www.dlib.org/dlib/July95/07weibel.html>.
- Willis, Craig, Jane Greenberg, and Hollie White. 2012. "Analysis and Synthesis of Metadata Goal for Scientific Data." *Journal of the American Society for Information Science and Technology* 63, no. 8: 1505–20.
- Zeng, Marcia Lei, and Lois Mai Chan. 2010. "Semantic Interoperability." In *Encyclopedia of Library and Information Science*, 3rd ed., 4645–62. New York: Taylor & Francis.
- Zimmerman, Ann. 2007. "Not by Metadata Alone: The Use of Diverse Forms of Knowledge to Locate Data for Reuse." *International Journal on Digital Libraries* 7, no. 1: 5–16.