THE DISSERTATION COMMITTEE FOR JULIA AMBER BULLARD CERTIFIES THAT THIS IS THE APPROVED VERSION OF THE FOLLOWING DISSERTATION:

CLASSIFICATION DESIGN: UNDERSTANDING THE DECISIONS BETWEEN THEORY AND CONSEQUENCE

Committee:

Diane Bailey, Supervisor

Melanie Feinberg, Co-Supervisor

James Howison

David Ribes

Karen Wickett

CLASSIFICATION DESIGN: UNDERSTANDING THE DECISIONS BETWEEN THEORY AND CONSEQUENCE

by

JULIA AMBER BULLARD, B.A. HONS., M.A., MASTER OF INFO. STUD.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

MAY 2017

CLASSIFICATION DESIGN: UNDERSTANDING THE DECISIONS BETWEEN THEORY AND CONSEQUENCE

Julia Amber Bullard, PhD

The University of Texas at Austin, 2017

Supervisor: Diane Bailey

Co-Supervisor: Melanie Feinberg

Classification systems are systems of terms and term relationships intended to sort and gather like concepts and documents. These systems are ubiquitous as the substrate of our interactions with library collections, retail websites, and bureaucracies. Through their design and impact, classification systems share with other technologies an unavoidable though often ignored relationship to politics, power, and authority (Fleischmann & Wallace, 2007). Despite concern among scholars that classification systems embody values and bias, there is little work examining how these qualities are built into a classification system. Specifically, we do not adequately understand classification construction, in which classification designers make decisions by applying classification theory to the specific context of a project (Park, 2008). If systems embody values— particularly values that might either cause harm (Berman, 1971) or provide an additional means of communicating the creator's position (Feinberg, 2007)— we must understand how and when the system takes on these qualities.

This dissertation bridges critical classification theory with design-oriented classification theory. Where critical classification theory is concerned with the outcomes of classification system design, design-oriented classification theory is concerned with

the correct processes by which to build a classification system. To connect the consequences of classification system design to designers' methods and intentions, I use the research lens of infrastructure studies, particularly infrastructural inversion (Star & Ruhleder, 1996) or making visible the work behind infrastructures such as classification systems. Accordingly, my research focuses on designers' decisions and rethinks our assumptions regarding the factors that classification designers consider in making their design decisions.

I adopted an ethnographic approach to the study of classification design that would make visible design decisions and designers' consideration of factors. Using this approach, I studied the daily design work of volunteer classification designers who maintain a curated folksonomy. Using the grounded theory method (Strauss & Corbin, 1998), I analyzed the designers' decisions. My analysis identified the implications of the designers' convergences and divergences from established classification methods for the character of the system and for the connection between classification theory and classification methods. I show how the factors—and the prioritization of factors—that these designers considered in making their decisions were consistent with the values and needs of the community. Therefore, I argue that classification designers have an important role in creating the values or bias of a classification system. In particular, designers' divergence from universal guidelines and designers' choices among sources of evidence represent opportunities to align a classification system to its community. I recommend that classification research focus on such instances of divergence and choice to understand the connection between classification design and the values of classification systems.

The Introduction motivates the problem space around values in classification systems and outlines my approach in focusing on classification design. The Literature

Review outlines the dominant theories in classification scholarship according to three elements of classification design: what decisions designers make, what information designers use in their decisions, and what skills designers apply to their decisions. In the Methods chapter, I introduce the site of my ethnographic research (The Fanwork Repository), detail my ethnographic methods, summarize the types of data I collected, and describe my grounded analysis. Three findings chapters examine one type of complex decision each: Names, Works, and Guidelines, respectively. In the fourth findings chapter, Synthesis, I define 10 factors designers considered across these complex design decisions. I then discuss how the factors figured into complex design decisions, how the factors overlapped and conflicted in design decisions, and how designers understood their role in making complex design decisions. In the Discussion chapter I connect the findings from the site of my ethnography to classification scholarship. In the Conclusion, I consider the contribution of examining classification systems as infrastructure, highlight the differences in accounts of classification design decisions made visible through classification theory and infrastructure studies approaches, and present suggestions for future research in classification design and the study of classification systems as infrastructure.

Table of Contents

x

List of Tables

List of Figures

Introduction

Classification systems are systems of terms and term relationships intended to sort and gather like concepts and documents. These systems are ubiquitous as the substrate in our interactions with library collections, retail websites, and bureaucracies. Classification systems make it possible to navigate large collections of documents or to order a domain of knowledge. Despite, or because of, the importance of classification systems to our access to and understanding of collected knowledge, a number of contemporary scholars have voiced concern that classification systems are causing harm (Adler & Tennis, 2013; Berman, 1971; Feinberg, 2007; Fox & Reece, 2012; Mai, 2010; Olson, 1998). In this tradition of critical classification theory, scholars argue that classifications cannot be objective and neutral; rather, classifications embody values and bias (Feinberg, 2007; Mai, 2010). Values, in this sense, might mean the narrow definition of personal and individual beliefs or the broader idea of "ethics" or the shared frameworks that govern behavior in a culture (Fleischmann & Wallace, 2007). For example, Olson (1998) argues that the Dewey Decimal Classification system's presentation of "labor" must be questioned as its definition excludes "unpaid labor" and therefore the collocation of books about it and the visibility of typically female-dominated forms of work. Despite changing cultural attitudes, the library shelves maintain an old-fashioned, patriarchal theory of what counts as "labor." Classifications systems, as our entry points into collections, impose these biases in subtle ways by directing our interactions and encounters with organized objects.

Through their design and impact, classification systems share with other technologies an unavoidable though often ignored relationship to politics, power, and authority (Fleischmann & Wallace, 2007). That classifications have these characteristics

is not the concern of only librarians and retrieval experts. To the extent that designers construct classification systems from particular cultural and social points of view, these systems can embody discriminatory views (Berman, 1971; Olson, 1998, 2000) and have real effects on the lived experiences of others (Bowker & Star, 1999). Several projects in classification theory have taken this premise as a starting point, identifying the particular bias of established classification systems (Bowker & Star, 1999) and suggesting revisions to ameliorate this bias (Berman, 1971; Kublik, Clevette, Ward, & Olson, 2003; Olson, 1998, 2000).

Scholars often state the position that classification systems embody values in contrast to early classification scholars such as Bliss (1929) and Ranganathan (1961, 1962) who claimed that classification systems can embody objectivity and fidelity to an external, real order. The claim that these systems are not purely logical or rational parallels the claims made regarding other types of "works" in science and technology, from scientific conclusions (Kuhn, 1964) to computer simulations (Galison, 1996). In reaction to the claim that classification systems must be built in ways that accept and recognize bias (Mai, 2010) some contemporary scholars argue that the impossibility of objectivity is overstated (Szostak, 2008). Regardless of the classification scholars' current relative certainty on the issue of value and bias, there is little work examining how these qualities are built into a classification scheme. Park (2008) notes that academic scholarship and instructional manuals on classification construction tend to focus on the epistemological basis or the mechanical details, respectively. We do not adequately understand the middle ground of classification construction, in which classification designers make decisions among terms and term relationships by applying classification theory to the specific context of a project (Park, 2008). If systems embody values— particularly values that might either cause harm (Berman, 1971) or provide an additional

means of communicating the creator's position (Feinberg, 2007)—it is important that we understand how and when the system takes on these qualities. An understanding of the links between classification design and classification system values can improve classification designers' awareness of how their personal and cultural values shape classification systems. For such an awareness to shape classification designers' work would require new approaches to classification design pedagogy, with an aim to avoid the presence of harmful bias in new and revised classification systems. Though few contemporary scholars would challenge the point that systems embody values, it is not clear at what stage of the design process value enters or how much agency classification designers have in shaping the social and cultural impacts of their projects.

This dissertation bridges critical classification theory scholarship with design-oriented classification theory. Where critical classification theory (e.g., Berman, 1971; Olson, 2001) is concerned with the outcomes of classification system design, design-oriented classification theory is concerned with the correct processes by which to build a classification system (e.g., Hidderley & Rafferty, 1997; Hjørland, 2013; López-Huertas, 1997). Therefore, this dissertation focuses on the middle ground of classification design, examining how designers practice classification design in order to trace back the consequences of classification system design to designers' methods and intentions. By reviewing classification research, I present the conflicting theories of classification design dominant in contemporary classification scholarship. By examining how classification theorists have defined the types of decisions classification designers make, the information classification designers apply to their decisions, and the skills relevant to classification design, I reveal the current state of our understanding of what role classification designers have in creating the character of our systems. This current state is a fractured one in which scholars begin from divergent theories of the purpose of

classification systems to advocate for different methods of classification design. For example, scholars who argue that the purpose of the classification system is to accurately represent reality advocate for classification design methods that feature scientific warrant, or the terms and term relationships experts agree are correct. . Research on classification design which connects classification methods to classification theory is sparse. Candid practitioner accounts (e.g., Wild, Giess, & McMahon, 2009; Young & Mandelstam, 2013) suggest that classification theory and classification design instruction present overly simplified visions of the classification designer's role. For example, Wild, Giess, and McMahon (2009) observed that the instructions on faceted classification leave out many details on how to make decisions and misleadingly present the faceted classification design method as straightforward and objective. Infrastructure studies presents an alternative research lens to illuminate classification design, particularly through infrastructural inversion (Star & Ruhleder, 1996) or making visible the work behind infrastructures such as classification systems. Accordingly, I present research questions that focus on classification systems change and rethink our assumptions regarding the factors that classification designers consider in making their design decisions.

In the methods chapter, I describe an ethnographic approach to the study of classification design that would make visible such factors and decisions. Ethnographic methods, including participant observation, diary studies, and interviews, provide a close view of classification designers making design decisions. This close view surfaces the exceptions and conflicts classification designers face and illuminates the middle ground between classification methods and classification theory. I took this approach to study the daily design work of a team of volunteer classification designers who maintain a curated folksonomy—a hybrid classification design approach in which designers adapt a

folksonomy consisting of user-generated tags into a controlled vocabulary system that accounts for synonyms, homonyms, and levels of specificity. Using a grounded theory method, I analyzed daily design decisions and found instances of classification designers reflecting on conflicting factors. I present these instances as three parallel accounts of designers' complex decisions.

Each of the first three findings chapters examines one type of complex decision: Names, Works, and Guidelines. Throughout each of these three findings chapters, I use designers' reflections on their design decisions to illustrate the factors designers considered when making complex design decisions. I go in depth to trace the dilemmas classification designers experienced, the factors that they considered in their decisions, and the effects these decisions had on the character of the classification system. In the fourth findings, Synthesis, I define 10 factors belonging to four types (definitional factors, user-centered factors, external truth factors, and technical context factors). I discuss how these factors figured into complex design decisions, how the factors overlapped and conflicted in design decisions, and how designers understood their role in making complex design decisions. Together, these four findings chapters represent classification design as it occurred at the site of my ethnography. Following the findings chapters, I return to the literature on classification theory and classification methods.

In the discussion chapter, I connect the findings from the site of my ethnography to classification research literature. I identify the implications of the designers' convergences and divergences from established classification methods for the character of the system these designers created and for the connection between classification theory and classification methods. I show how the factors—and the prioritization of factors— that these designers considered were consistent with the values and needs of the designers' community. Specifically, the designers' feminist and creative community of

fanwork creators valued a multiplicity of perspectives and prioritized the minimization of harm over the efficiency of a retrieval system. Therefore, I argue that classification designers have an important role in creating the values or bias of a classification system. In particular, classification designers' divergence from universal guidelines and designers' choices among different sources of evidence represent opportunities to align a classification system to its community. I show how reflection on complex design decisions reveals the theory behind classification design. I therefore recommend that classification research continues to focus on such instances of divergence and choice to understand the connection between classification theory and classification method.

In the conclusion, I consider the contribution of examining classification systems as infrastructure and highlight the differences in accounts of classification design decisions made visible through classification theory and infrastructure studies approaches. Finally, I present suggestions for future research in classification design and the study of classification systems as infrastructure. Classification systems will continue to play an important role in how we interact with and understand collections. How well we understand the role of classification designers in determining the character of these systems will shape how we instruct future classification designers and how we interpret the systems we encounter in daily life.

Classification systems play an important role in facilitating our access to collections. In simplifying a large set of documents or concepts into sets of similarity and relatedness, these systems of terms and term relationships instantiate particular values and biases. Especially as these systems fade into the background of our interactions with collections, classification systems can impose and advance a value or bias such as patriarchy. As a starting point for my research, I challenge this invisibility by focusing on classification design; classification systems—their values and bias—are not inevitable but

are the result of a series of decisions. In the following chapter, I turn to classification research to ask: what do we know about classification system design?

Literature Review

This literature review explores the field of classification research with a particular focus on accounts of classification system design. For the purposes of this review, I define "classification systems" broadly, referring to all designed systems of term relationships intended to sort and gather like items. I include in my review representative studies in contemporary classification research as well as the early texts that provided the foundational arguments for classification theory's dominant schools of thought. What we know about the components of classification design from these studies and texts is the foundation for our understanding of how classification systems come to take on values and biases, such as those of their designers and their designers' cultural contexts.

Unfortunately, the scholarship of classification research has rarely examined classification design in detail, and hence has left unexplored the processes by which values and biases arise in classification systems. The exceptions are works that undertook detailed ethnographic study (Park, 2008) or works in which classification designers provided candid accounts of their own work (Kublik et al., 2003; Wild et al., 2009; Young & Mandelstam, 2013), which I feature in this review. Thus, I examine the scholarship on classification research that most closely approached studies of classification design: those studies that examine classification as a practice rather than as an encountered object or technology. Because most classification research describes design at arm's length, my review relies on proxies found in the literature representing elements of classification design. For example, I use the field's discussion of valid

evidence to represent the element of designers' information needs and the competing conceptual approaches to design to represent the elements of the skills scholars believe classification designers bring to bear on design decisions.

In this review, I present the elements of design as the types of decisions classification designers make in the course of their work, the information they apply as evidence or authority for their decisions, and the skills theorists argue are at the core of this practice. For each section, I categorize the findings from classification research, highlighting the points of consensus and dispute within this field. For example, in the section on the information of classification design, I categorize the classification research literature according to four types of "warrant"—sources of evidence designers use to derive or justify terms and term relationships. In categorizing the findings from classification research, I considered established definitions from the field, such as Beghtol's (1986) types of warrant and Mai's (2005) distinction between document and domain-oriented approaches. My categorizations are a few among many possible lenses on classification research; my aim in employing them is to recognize the major points of view in the field, their shared tensions, and their limitations in describing how classification design accounts for the character of classification systems.

CLASSIFICATION DESIGN DECISIONS

In creating a classification system, classification designers make a series of decisions based on the information available and the designers' assumptions regarding the goals and potentials of classification systems. In later sections, I describe how

classification research literature defines the information relevant to decision making and the skills the classification designer applies to the project. In this section, I review how the field of classification research defines and addresses the decisions themselves: to what extent scholarship in this field explores each facet of classification design, and what concerns scholars pose regarding the outcome of these decisions.

A great deal of classification scholarship has been practitioners' accounts of their work on new projects and revisions to existing systems. Together with the conceptual literature that has explored the rationale and outcomes for making particular decisions, practitioners' accounts have revealed the field's focus on different steps in classification system design and maintenance. In the following sections, I describe the field's uneven division of focus between issues of syntax and semantics, identify the major areas of concern within each of these categories, and, finally, pose the argument that semantics remains the less explored and more engaging realm of decision making, particularly when pursuing enduring questions of values in classification systems.

Choices of Syntax and Choices of Semantics

The types of decisions made by classification designers can be grouped into two types: the choices of syntax, or the rules for the format and relations between elements in the system, and the choices of semantics, or the meaning of elements in the system. Classification scholars adopted these terms from linguistic theory, in which syntax is the realm of pure grammar and semantics of meaning in words and sentences (Friedman & Thellefsen, 2011). For example, in classification construction, syntax might determine

whether countries of the world appear in one flat category or hierarchically below continents or colonial empires, while a semantic decision would determine whether Taiwan counts as a country.

Jones (1970) summarized the major syntactical divides by categorizing classification methods according to decisions on how to structure three types of relations: those between properties and classes, those between objects and classes, and those among classes. The field of classification research has kept a steady debate on the merits of each of these choices and their expression in notation (Clare Beghtol, 1994, 1995; Broughton, 2006; Mills, 1964; Ranganathan, 1962; Soergel, 1974). Arguing for a focus on semantics, Beghtol (1986) argued that classification research had been overly focused on syntax, singling out the work of the Classification Research Group (CRG) as particularly preoccupied with perfecting rules for notation and the relationship between terms (pp. 109-110). The CRG, which produced specialized classifications for individual fields of knowledge, still focused on the ideal syntactical expression of a domain with less focus on the meaning of terms in that field (Foskett, 1974; Langridge, 1976; Vickery, 1964). This trend continued in some contemporary classification approaches, such as the Facet Analytical Theory in Managing Knowledge Structures for Humanities (FATKS) project, which explored the potential for faceted syntax to facilitate machine handling of vocabulary maintenance (Broughton & Slavic, 2007). Discussing the meaning of classification systems, Beghtol argued, always requires a discussion of semantic decisions, though this discussion is less explicit in classification theory (1986).

Choices of syntax, which determine the possible ties between elements in the classification system, tend to change as our technology changes (López-Huertas, 2008; Mills, 2004; Smiraglia, 2002). For example, while early classification scholars assumed a physical library and therefore a single shelf order (Cutter, 1891; Dewey, 1876; Ranganathan, 1962), electronic catalogs of digital surrogates and digital materials without a stable retrieval order suggested different solutions (Albrechtsen & Jacob, 1998; Broughton, 2006). Changes in theories of semantic choices, however, have followed changes in our understanding of language and meaning, with major shifts following the work of Wittgenstein (1958), Rosch and Mervis (1975), and Lakoff (1987). For example, the field's approach to problems of aboutness (Furner, 2004; Hjørland, 1992) has reflected a broader shift in theories of language in meaning from an Aristotelian ideal of external, mutually exclusive categories to the fuzzy systems of family resemblances, prototypical examples, and the influence of culture and lived experience on our interpretations of sameness and difference. In the following two sections, I summarize contemporary classification research on syntax and semantics.

Choices of Syntax: Structure, Schemes, & Indexing Rules

Among the most important choices of syntax the classification designer makes is the type of structure, scheme, or rule set to be used for the project as a whole. The classification designer—assuming he or she is given perfect autonomy regarding the classification system—decides amongst competing structures such as taxonomies, thesauri, or subject headings. While textbooks in classification have provided definitions and practical considerations for each of these structures, the field of classification

research has often been concerned with comparing the theoretical basis of competing genres of organization (Friedman & Thellefsen, 2011; Jacob, 2004) or testing the genres' compatibility and interoperability in live systems (Wang, Chaudhry, & Khoo, 2008; Yi & Mai Chan, 2009).

Part of the difficulty in differentiating among the characteristics of the systems such as thesauri and ontologies has been that each has a colloquial meaning apart from its specific use in information organization (Gilchrist, 2003). Some scholars have argued that discussing these systems without acknowledgement of their differences impedes advances in classification theory and practice (Jacob, 2004). Even when the meanings and differences are clear, the preferred system for a given knowledge organization problem is unclear. Based on theoretical and technological changes, scholars have argued for the primacy of one system over another, often citing predicted performance in retrieval (Broughton, 2006; Mills, 2004; Munk & Mørk, 2007). Others have explored the possible compatibility between systems, seeking methods to translate one system to another or—particularly with the introduction of folksonomies—exploring how one system might augment another (Yi & Mai Chan, 2009).

Once classification designers choose a type of system, they can decide how strictly to adhere to its traditional structure. For example, Wild, Giess, and McMahon (2009) reported that the instructions for a system such as faceted classification still left some choices unstated, particularly the parameters given to classifiers when applying terms to individual documents. In large projects, in which the classification designer and classifier are distinct roles, part of the classification designer's task is to determine the

rules guiding the classifier or cataloguer's work; removed from the day-to-day reality of providing subject descriptions to documents, the classification designer must rely on theoretical principles and imagined practical outcomes to decide these rules. The classification designer must also determine how the system might be edited in the future; while Soergel (1974) gave detailed instructions on creating a maintenance schedule, instructional textbooks have rarely included maintenance as an element of classification design. This absence has persisted despite the impact of policies such as revision procedures on how adaptable these systems can be to future cultural and practical realities (Kublik et al., 2003).

Choices of Semantics: Meaning, Term Equivalence, and Hierarchies

Scholars have often argued the merits of different syntactical choices through theoretical principles and imagined impacts on retrieval. Decisions of meaning, however, which characterize semantics, are best demonstrated through case studies describing classification designers' processes in arriving at particular terminology for their system. In designers' accounts, scholars have described in detail decisions on term equivalence—which terms they defined as being interchangeable and which among synonyms they privileged in the system (Clare Beghtol, 1995; Ibekwe-SanJuan, 2006; López-Huertas, 1997; Nielsen, 2001; Shearer, 2004). Some scholars have argued for exhaustive approaches to term equivalence decisions, seeking to provide the greatest number of possible entry points for users with diverse vocabularies (Bates, 1986; López-Huertas, 1997).

Discussions of meaning are also central to choices of relationships between terms; while the possible relationships between terms are decisions of the system's structure and hence an issue of syntax, how subjects are ordered and built into hierarchies is the realm of semantics. Scholars have tended to be particularly concerned with how the top level terms of a hierarchy are decided: whether they correspond to the most central and stable concepts of a field (Clare Beghtol, 1995), delineate the major divisions of a field (Hjørland, 1998), or represent fundamental categories (Mills, 2004). The interest in top level terms has likely been due to the importance these terms have in shaping the system as well as the relative simplicity in examining such a limited number of terms. Olson (1998) took a different approach, examining in detail lower levels of hierarchy to describe the interaction of many definitions on the location of a single term. For example, Olson revealed the impossibility of placing "unpaid labor" under the DDC "labor" divisions due to the latter's definition, and explored what changes are necessary to fit new concepts into such a system (pp. 248-251).

Summary

Decisions of structure (syntax) and decisions of meaning (semantics) have been the explicit subjects of case studies in classification research because case studies document the rationale and process at arriving at particular systems. However, these decisions have also been at the center of conceptual works in classification research, as scholars have debated the impact and importance of different types of decisions and the correct way of approaching decisions such as the shape of the system or the ideal

characteristics of top level terminology. This section has briefly reviewed how the field of classification research has described the decisions made in classification research.

Scholars who have reviewed classification research at the abstract level of types of decisions have argued that the field suffers from an excessive focus on matters of syntax over semantics (Beghtol, 1986) and have linked developments in syntax to changes in technology (Smiraglia, 2002; Mills, 2004; López-Huertas, 2008) while noting that semantics tends to change at the slower pace of conceptions of language (following Lakoff, 1987; Rosch & Mervis, 1975; Wittgenstein, 1958). In discussions of fairness, representation, and cultural values, scholars most often invoke decisions of semantics (Fox & Reece, 2012; Kublik et al., 2003; Olson, 1998). For these reasons, in the remainder of the literature review I draw from examples of semantic, rather than syntactical, design choices in classification systems.

Scholars in classification research have described two necessary inputs to classification design: the information the classification designer requires to complete the system and the skills the classification designer brings to bear on the project. The information the classification designer requires can be described as types of warrant, or the sources and forms of evidence used as the raw material for terms and term relationships.

THE INFORMATION OF CLASSIFICATION DESIGN: TYPES OF WARRANT

Warrant provides the basis for making decisions on which terms represent which referents as well as how terms relate. Beghtol (1986) provided a review of classification

theory from this perspective, examining how different scholars have interpreted core classification principles such as literary warrant. Beghtol defined warrant as "the authority a classificationist invokes first to justify and subsequently to verify decisions" about terms, their relationships, and their order (p. 110). When claiming adherence to a type of warrant, classification designers identify the primary form of evidence used in their decision making. Classification theory has associated each type of warrant with a history of arguments regarding the purpose and potential of classification systems, which I highlight in this review. In the following sections, I outline the major principles of warrant—literary, scientific, user, and ethical—as the information needs of classification design. I choose these types of warrant not only because they align with dominant categorizations of classification principles (Clare Beghtol, 1986; Hjørland, 2013a) but also because these types of warrant represent the strongly held and mutually contradictory positions persistent in classification research.

Literary Warrant

Literary warrant, which Hulme (1911) first articulated as a principle of classification design (1911), is a component of most classification systems, though the mechanics vary widely from the themes arising from the literature represented in the library collection to the terminology used in the library holdings (Beghtol, 1986). The basic premise of literary warrant is that classification designers derive terminology from the scholarship they are organizing. For example, Hulme (1911) argued that a chemistry library would not need "Gold" as a subject heading since no monographs on individual

elements had been published (p. 46). Literary warrant has been a particularly appropriate principle for systems that seek to represent works as they present themselves. Literary warrant has been central to the domain analysis approach, in which the classification designer examines the paradigms of the field to be organized and creates a system that represents these works according to the domain's own logic (Hjørland & Albrechtsen, 1995).

Literary warrant can inform the design of a classification system at different levels. The idea of "a literature" (Agre, 1995) might be the starting point for defining the scope of the system, as in domain analysis (Hjørland & Albrechtsen, 1995); may provide the terminology to name classes and facets in the system (Vickery, 1960); or may be the concern of the classifier and cataloguer when individual works are integrated into the system (Ranganathan, 1962). Where literary warrant has been applied exclusively, Mai (2005) and Fidel (1994) referred to the method as the document-centered approach in which the work itself is examined without context. Mai (2005) described the document-centered approach as an unrealistic goal, given that the classification designer or indexer always uses other information to make judgments regarding appropriate terms, such as subject knowledge of the field as a whole and imagined user needs; the assertion that the classification worker can determine subjects purely from the document itself is a stance that has never borne out in practice. Pretending otherwise merely conceals the classification designer's interpretive work (Mai, 2005). In contrast, domain-centered approaches have intentionally integrated this kind of contextual information and typically

complemented the information from the works themselves with scientific and user warrant (Beghtol, 1986; Mai, 2005).

Scientific or Consensus Warrant

Bliss (1929) replaced literary warrant as a central principle of evidence with scientific or consensus warrant, organizing knowledge and individual works according to the contemporary scientific conclusions on the relation between the fields and the nature of knowledge. Rather than let the literature speak for itself, the organization occasioned by scientific warrant positioned each field as subject to the logic of scientific fields with greater claims to the organization of knowledge, so that biology informed the location of psychology and physics informs the location of all natural sciences (Bliss, 1929, pp. 262-277). Unlike the example from Hulme, a system organized according to scientific warrant would include an entry for Gold as it is a recognized concept within the field of chemistry. Bliss presented this approach as compatible with a universal, hierarchical system in which all works and types of knowledge must have a place; relying on each field of work to define its own place in relation to others in such a system would create contradictions and inconsistencies. For example, spiritual works and scientific works would contradict which has the greater claim to knowledge. Relying on the consensus of the scientific view from one perspective would provide the possibility of arranging each field in relation to one view of knowledge.

Though it has not always been invoked by name, Bliss' scientific warrant is as ubiquitous as literary warrant in classification research, sometimes appearing as

consensus warrant. Beghtol (1995) noted that most classification design guides include references to consulting subject experts and she argued that the idea of arranging a field according to the consensus of its experts is a continuation of Bliss' principle. The variation here—similar to the use of literary warrant—is in the universal to the particular; are the subject experts meant to define the main classes of the classification system in line with their philosophy of knowledge, or to assist the classification designer with interpreting the divisions within a field and the jargon in its works? Domain analysis has been among the contemporary approaches that explicitly prioritize scientific warrant; along with information from published works, experts' understanding of their field of knowledge is at the core of how the classification designer organizes a literature (Hjørland & Albrechtsen, 1995).

User Warrant

User warrant has had a shorter history as an explicit source of evidence for classification decisions, but has been a constant thread in the library and knowledge organization literature. A defining work of library science, Ranganathan's (1964) *Five Laws of Library Science*, oriented all library services towards the user. Ranganathan (1962) applied this orientation directly to classification through the principles of helpful sequence, a series of guidelines determining the order subjects should occur in order to match the expectations and logic of readers. Indeed, one of Mai's (2005) criticisms against the document-oriented, literary warrant approach was that the imagined user is always used to support the classification designers' decisions.

The turn toward user-oriented library services and system design has introduced new terminology and methodologies to this approach in a reaction to a perceived system-centric paradigm. Albrechtsen and Jacob (1998) attributed this shift to changing library technology, which provided the possibility for direct user access to classification systems without reliance on the librarian as intermediary. In the computer-supported, automated library, the user is seen as autonomous, and classification research is tasked with creating systems that a novice user, rather than an expert librarian, can use for subject access (Albrechtsen & Jacob, 1998). The user-oriented paradigm in library services and its corresponding classification research prioritized user information as the most important element in system construction. Classification designers used search strings entered into electronic forms to learn what terminology users expect will produce results (Soergel, 1974; Fidel, 1994), surveys to map out users' mental models of the literature they seek (López-Huertas, 1997), and observation protocols to determine the ways users interact with documents (Albrechtsen & Pejtersen, 2003).

Bates' (1986) end-user thesaurus represented an early approach to offering greater autonomy to the user by creating a classification system tailored to user vocabulary. Bates proposed creating an end-user thesaurus distinguished from the traditional thesaurus by its extensive entry vocabulary referring users to the system's preferred terminology. By including every possible variation of names for a concept, including slang terms, Bates hoped to bridge the natural language of users with the controlled vocabulary of library. While this approach provided an additional entry layer to address user needs, more recent classification systems have attempted to integrate user language

into the core vocabulary itself (López-Huertas, 1997). In the first approach, the core system was augmented to serve user needs; in the latter, the basis of the core system was the user's perspective.

The division among classification designers who prioritize user warrant—and a criticism often levied from designers who do not—has been whether there is an expectation of a single user need that might be derived from search terms and user studies (Feinberg, 2007; Hjørland, 2013a). In cognitive approaches such as López-Huertas' (1997), the intent has been to identify the mental model of scholars in a field, that of users, and use the system as a point of translation and inscription through which users can retrieve documents while learning the structure of the domain. The information the classification designer used in this method is an aggregate summary of user studies representing a generic user contrasted against an expert author (López-Huertas, 1997). This assumption of a singular user perspective seemed particularly inappropriate for large systems distributed among a diverse user community (Mai, 2010).

A related criticism against basing a system on user warrant has been the limited potential of such a system. Bliss (1929) considered and dismissed adapting classifications to user expectations because users would be better served by adapting and learning classifications based on scientific knowledge. Hjørland (2005, 2013) repeatedly questioned user warrant on the grounds that users do not know what they do not know; creating classifications based on users' limited and flawed understanding of a field can only create limited and flawed systems. This criticism has been consistent with classification theorists who viewed systems as cognitive scaffolding, and stressed the

impact that classifications have not only as retrieval methods, but as teaching devices (Jacob, 2001). For these scholars, designer operating according to user warrant would achieve a system with immediate ease of use while sacrificing the educational potential of the system.

Ethical Warrant

The turn toward recognition of both the subjectivity of classification design and the impacts of systems on users and collections has spurred the urgency to consider ethical warrant, the notion that the final justification for decisions of terminology and structure must be ethically sound. To consider ethical warrant as an input to the system means interrogating the bias and impact of all other inputs: Does the collection use outdated and discriminatory language? Is the diversity of our user community represented by our search logs? Do our aggregate methods of analysis obscure minority concerns? How do we identify experts, and should we address divergence from the consensus view?

Ethical warrant has been less a material for classification design than a position on the responsibility of the designer in applying any type of warrant. Fox and Reece (2012) demonstrated that classification designers can act according to one of many ethical frameworks, each of which produces a very different kind of classification system, from a Kantian or Rawlsian ethic that prioritizes universal neutrality over local concerns to an ethic of care that considers context and allows for exceptions and creativity.

Particularly for large and universal classification systems, Beghtol (2002) argued that sensitivity to ethical concerns such as political, religious, cultural, gender, and

language representation is a vital part of classification design. To the extent that the collections and users of classification systems are increasingly international and diverse, consensus warrant is more difficult to apply and more likely to privilege the viewpoint of one set of users at the expense of another. Unlike the previous types of warrant, an ethical warrant is less likely to provide a source of vocabulary than it is to provide evaluative techniques to consider the impact of particular classification design decisions. For example, an inclusive design that aims to achieve the ethical value of cultural hospitality would choose to represent time in multiple formats, rather than privilege a single culture's representation of the concept (Clare Beghtol, 2002).

Summary

Which type of evidence, or which warrant, to use in justifying and verifying classification decisions is a choice tied to the classification designer's perspective on the purpose and potential of classification systems. It is not a decision made separately from other major design decisions such as the scope of the subject matter to be organized or the type of syntax to be applied. Each type of warrant implies particular assumptions about how users interact with a system, how documents contain subjects, and whether a classification system can teach as well as retrieve. By dividing classification scholarship into the categories of literary, scientific, user, and ethical warrant, I have shown how the core assumptions regarding the purpose of classification determine the sources a classification designer is likely to consult, and how traditions of teaching classification

38

designers to pursue certain types of evidence are likely to produce classification systems with different character and value.

The major arguments within classification theory over types of warrant have been of two types: whether a given type of evidence can provide the type of guidance classification designers need in their work and whether the result is a good classification system. Mai (2010) and Fidel's (1994) criticisms of literary warrant fell into the first category; they questioned whether any classification designer could avoid using information from outside the text itself, despite ostensible adherence to this principle. Hjørland (2013) and Bliss's (1929) criticisms of user warrant fell into the second category; they were familiar with classification systems built from empirical evidence of users' vocabulary, expectations, and search behaviors, but believed that systems built from these data reinforced naïve and false conceptions of fields of knowledge. Neither objection has been resolved within the existing literature. Scholars championing the pure application of any warrant have not provided accounts of classification designers who rely purely on that source of information and arguments regarding the long-term effect of types of warrant on users have remained conceptual rather than empirically based.

Though scholars have presented warrant as the sole input into the terminology of a classification system, other factors are likely present, even if only because designers will encounter incomplete or conflicting answers from types of warrant (Mai, 2005). Specifically, the designers' judgment is likely to play a role if a factor such as warrant is not adequate or clear. In the following section, I shift from the material designers use to how designers make decisions. I review how classification literature defines the skills of

classification designers to answer the following question: What is the role of classification designers in making the decisions that determine the character of a classification system?

THE SKILLS OF CLASSIFICATION DESIGN

How the classification designer goes about making decisions in creating and maintaining systems is a process on which the field of classification research has been highly divided. While informal discourse often characterizes categorizing things or concepts into classes as a natural and unproblematic activity, scholars in classification research have described the process of creating a system of such classes for others' use in complex and contradictory ways. Classification research has complicated the natural and unproblematic image of categorizing things by portraying the classification designer as engaging in rule following or creativity, beginning from universal principles or local context, and acting autonomously or in negotiation with other workers. In the following sections, I categorize the versions of classification skills or ideal approaches to represent the active and influential points of view in the field of classification research. In particular, I aim to highlight their points of conflict with regard to principles and practice. As these positions coexist in contemporary scholarship, they are not only historical developments in the field's view of classification practice but also competing visions of how classification designers conduct—or should conduct—their work. Each of the positions complicated and argued against the foundational or "folk theory" (Mai, 2010, pp. 631-635) of classification: that classification designers arrive at correct systems

through a systematic though straightforward process of grouping things which are inherently similar.

Therefore, this section of the literature review outlines the major schools of thought in classification research regarding the nature of classification design: classification as human nature, classification as rule following, classification as application of domain expertise, classification as teamwork, and classification as a personal, creative act. I examine the major point of contention—to what degree rules and standards decide classification design, and to what degree the classification designer determines the outcome—alongside less overt debates, such as whether the classification designer works alone or shares decision making tasks with coworkers of different professions. Each of these positions has assumed a particular goal or potential for classification systems, a particular type of classification designer, and a different level of autonomy in the work. By framing these different positions in terms of the classification designers' skills, I connect the designers' decisions and actions to classification theory. Each of these sections provides a different answer to the question: What skills does the designer apply to classification design? Therefore, each section also provides a different answer to a second question: What is the role of the designer in determining the character of a classification system?

Classification as Human Nature

Classification design has sometimes been portrayed as an extension of the typical human practice of organizing experience and knowledge into systems of classes and

relationships. Classification systems, in this view, come about through a more rigorous application of our natural processes of dividing groups of things. The subject of Bliss' (1929) *The Organization of Knowledges and the System of the Sciences* was the development of a specific classification system for subjects in science. In the early chapters, Bliss defined organizing—and organizing knowledge in particular—as a ubiquitous social act and a foundational mental act. Bliss framed the classification of the sciences as one part of humanity's project of organizing society for the betterment of all, and connected the project of creating classification systems to the progression of human society through organizations such as labor unions and the League of Nations. This argument was not specific to Bliss; his contemporary and the author of the Universal Decimal Classification, Otlet (1990), made similar connections between the practice of classification and the trend toward global organization. Arguing society progresses towards greater organization, Bliss provided a historical overview of the rise of bureaucracies, lobby groups, philanthropic organizations, and professional associations. For Bliss (1929), these social organizations were also engaged in the organization of knowledge, and progress in the organization of knowledge could inform more cooperative and more productive organization of society.

The everyday character of systematic organization, for Bliss, was a positive aspect of human nature and an integral part of human progress. Writing in the aftermath of World War I, Bliss went as far as to state that a more organized human society, informed by more organized knowledge, would be cooperative and productive enough not to engage in such violent conflict. On the other side of this perspective have been scholars

such as Zerubavel and Olson. In *The Fine Line: Making Distinctions in Everyday Life*, Zerubavel (1991) stated that organizing our knowledge and perceptions according to sameness and difference is at the core of how we understand the world around us, but that these heuristic methods are not uniformly positive, and are also at the center of stereotyping and racism that lead to atrocities such as the Holocaust. Olson (2001) argued that while sameness and difference are ingrained ways of organizing the world, they are less natural than learned practices belonging to a particularly Western cultural perspective. The major difference between Bliss's position and Zerubavel and Olson's positions on the everyday nature of classification is that Bliss framed it as progress; though we may sometimes be wrong, consensus should always lead towards classification systems that more accurately reflect reality. Zerubavel and Olson also saw this organizing as a common personal and social tool, but noted how organizing our experience according to sameness and difference also leads to the very types of tragedies Bliss promised it would prevent. Which version of classification-as-human-nature a scholar follows is likely to determine the extent to which they temper these instincts with rules and external information. In one version, classification is an intuitive skill the designer develops and refines; in the other, it is an intuitive skill the designer must challenge and second guess.

Whereas classification-as-human-nature posits that classification is a skill that belongs to us all, the remaining sections posit classification as a more intentionally developed skill among a few. The order of sections follows a progression that roughly matches the chronological development of classification theory, though each of these

perspectives remains influential in contemporary classification research. The order begins at one extreme (that the skill of classification is in following rules)to its opposite (that classification is a creative and personal skill).

Classification as Rule Following

Some scholars in classification have presented classification design as the unquestioned application of rules. Rules produce a system that is logically correct (Ranganathan, 1962), that supports interoperability between institutions (Hoffman, 2009; Theimer, 2012), and that satisfies certain ethical perspectives (Fox & Reece, 2012). The primary skill of classification for many scholars was therefore an adherence to and respect for the rules of classification system construction. As with defining classification as the extension of human nature, classification design as rule following is not an exclusive profession but a type of task that would not suit all interests. For scholars who prioritized rule following as a classification skill, the standards and steps of design that instructional works provided to the classification designer were the most important form of information and determined all of the choices in the process. This perspective has been seemingly implicit in any instructional manual or theoretical account that does not mention creativity or the application of judgment to subjective choices (e.g., Aitchison, Gilchrist, & Bawden, 2000). The dominant view in these contemporary instructional texts on classification design has been the application of a proven process. Hoffman (2009) highlighted the importance of adhering to rules in works regarding systems to facilitate centralized organization and coordination among institutions. In these cases, the

expectation of the classification community has been that all classification designers will arrive at the same conclusions, and therefore create a consistent set of subjects and indexed documents that can be easily exchanged across different local collections (Theimer, 2012).

Ranganathan's (1962) guide to faceted classification for libraries exemplified the centrality of rules to classification design—all design decisions are subject to a single principle. While Ranganathan presented classification design as common sense, in his instructions to classification designers the focus was on the principles and rules that produce consistent and intuitive systems. Through *The Elements of Library Classification*, Ranganathan (1962) guided the classification designer through a series of exercises that illustrate the helpful sequence, the one ideal order of facets which the library user would find most intuitive and useful for locating books by subject. This single order was accomplished by following the 33 canons and 13 principles Ranganathan derived for the reader in the course of the text. For example, the Principle of Later-in-Evolution mandated the ordering of classes according to their development in reality, so that Elementary Education precedes Secondary Education, while the Canon of Decreasing Extension placed the general before the specific, so that Education precedes Elementary Education:

> [...] it seems reasonable to give precedence to Elementary Education on the
>
> grounds that it belongs to an earlier stage of educational evolution [...] The

principle we have used by which Elementary education has been given precedence over Secondary education is the one stated below.

44. Principle of Later in Evolution

If two classes belong to two stages in the same line of evolution, the one at the earlier stage must have precedence over the other.

(Ranganathan, 1962, pp. 33-34, emphasis in original)

Ranganathan characterized certain rules as derived from common sense, while others required an arbitrary choice between two equally valid orders. In his "Canon of Consistent Sequence" Ranganathan made clear that even these arbitrary choices must be followed throughout the system: "When the same categories occur explicitly or implicitly in different places, they must be arranged in the same sequence in all the places" (p. 36), so that if the order of "Indian Law" before "British Law" was decided arbitrarily, Asian countries must always precede European countries in other lists. Ranganathan also provided rules for possible conflict between rules, providing the classification designer with a hierarchy of his canons and principles from which any contradictions in order could be resolved, and resolved consistently across all classification designers.

Ranganathan's focus on the centrality of rules (or canons, or principles) to classification design was consistent with his characterization of the personality of the ideal classification designer. Ranganathan described the librarian who follows his method as one with more will to pursue detailed work than the "lazy librarians" who did not appreciate its function (p. 10). To understand the helpful sequence only required the

application of logic and common sense; to follow it in creating classification systems and classifying individual documents was only a matter of checking boxes in regards to all the potentially relevant canons and principles Ranganathan provided. The result was a process of translation, but unlike literary translation, it involved no creative interpretations; it was "a steady one-way process which can have only one conventionally perfect result" (p. 146). In fact, for Ranganathan, any special perspective the librarian has should be left out of the classification, including subjective matters such as opinion. As with the exercises that provided the framework of Ranganathan's lectures, the goal was for each librarian to produce the same classification, with any differences being the correctable products of a lack of experience. This characterization can be contrasted against Ranganathan's characterization of reference work, in which the librarian's intuition and "flair" came to play in their interaction with library patrons (Ranganathan, 1961, pp. 53, 73-75).

The level of confidence in the completeness of rules and their ability to solve all dilemmas Ranganathan's guide expressed is remarkable. This impression has more to do with Ranganathan's (1962) conversational style, in which he repeatedly used phrase such as "a moment's consideration will remind them" (p. 40) to underline how elementary he found his logic, than a difference between the importance of rules in that text and contemporary texts such as *Thesaurus Construction and Use: A Practical Manual* (Aitchison et al., 2000). Current editions of classification constructions manuals have not repeatedly admonished the reader to appreciate the simplicity of the rules; rather, manuals have presented the rules as a complete set of information that will produce a correct

system. Theimer (2012) noted that the field's focus on rules and standards has a practical purpose: the facilitation of inter-organization cooperation and coordination of systems. In Theimer's view, textbooks and supervisors instructed classification designers to follow the given rules not because individual judgment and creativity were deemed absent, but because these skills were idiosyncratic and would lead to differences in design decisions.

Having everyone follow the same rules satisfies an ethical need, as well as a practical one. In their review of the ethical frameworks behind various approaches to classification work, Fox and Reece (2012) identified the American Library Association's modus operandi in its approaches to designing systems and training workers as appealing to the Rawlsian "veil of ignorance" (p. 380). Rawls' justice ethic guides decisions by asking, "What system would I design if I did not know my own status within that system?" While useful for creating a singular, universal system that does not disproportionately benefit those in power, its instantiation in a formal set of rules removed all local control. Classification designers then became a means to enact standards, rather than having any decision making authority based on their insight into the specific characteristics of their system and its users (Fox & Reece, 2012). This perspective most obviously clashed with the definition of classification design as creative work but it also contradicted the idea that classification designers require domain knowledge to fulfill their tasks. If the rules alone can produce a working system, the specific knowledge of the classification designer was unnecessary and even ignored in favor of the universalist position.

Taken to its logical extreme, the position that classification is a procedure of following set rules leads to automated classification, in which the work of determining the terms and relationships in a classification system is delegated to computer algorithms. The goal of creating automated systems that can perform this work has been a strong thread in classification research since the widespread use of computers in libraries (Spärck Jones, 1970) but has grown quickly with the introduction of the internet and the rapid growth of document collections (Golub, 2006; Svenonius, 2000). Researchers have developed techniques such as machine learning (Labrou & Finin, 1999; McCallum, Nigam, Rennie, & Seymore, n.d.; Sebastiani, 2002) and document clustering (Cutting, Karger, Pedersen, & Tukey, 1992; Jain, Murty, & Flynn, 1999; Wacholder, Evans, & Klavans, 2001) to create subject terminology and classificatory structures as well to apply subject terms to documents. Though these methods typically have begun with a large set of documents to produce controlled vocabularies, in a reversal of the typical top-down process of classification design, they have shared with the traditional rule-based approach a confidence in the inherent, objective subject matter of documents and the universality of subject descriptions.

Scholars in classification research have critiqued this largely information retrieval-oriented approach. In a review of automated classification approaches, Golub (2006) concluded that the common assumption of the various approaches is that "concepts have names," which caused trouble in humanities and social sciences domains (p. 363). Golub also argued that common performance measures, which included information retrieval metrics of precision and recall, and indexing consistency with

experts and end users, had theoretical flaws and relied on a simplistic understanding of how users recognize subjects. Svenonius (2000) questioned the foundation of these methods, arguing that algorithms fall short of the "truly intellectual" tasks of determining meaning and significance (p. 198). These classification scholars questioned the extent to which automated methods—and strict rule-based methods broadly—could replace the human work of semantic judgments.

Classification as Application of Domain Expertise

The common core of classification designers' knowledge has been an understanding of the rules, procedures, and principles of classification; some classification designers have differentiated themselves from each other and from computer algorithms through their domain expertise in the bodies of knowledge to be organized. The importance of domain knowledge has been particularly clear in classification case studies that highlighted the local (Kublik et al., 2003), including concerns over marginalized bodies of knowledge (Nero, 2006; Olson, 1998) and scholarly domains that do not match the structure of dominant fields (Foskett, 1974). Marking a transition from the universal classification systems of Dewey, Bliss, and Ranganathan, the Classification Research Group (CRG) championed an approach that tailored the construction of systems to the structures and needs of individual domains of knowledge or work (Foskett, 1974; Langridge, 1976; Vickery, 1960, 1964). More recently, Hjørland and Albrechtsen (1995) presented the domain analysis approach as the ideal method to produce systems that represent the specific and multiple points of view

within a knowledge domain. Similarly, the cognitive work analysis framework provided methods to tailor a classification system to a specific work environment (Albrechtsen & Pejtersen, 2003; Fidel & Pejtersen, 2004).

The goal of the CRG was twofold: to further develop the faceted classification methods introduced by Ranganathan and to produce classification systems tailored to the structures and needs of specific fields of knowledge. Various members of the CRG produced detailed accounts of the rationale behind and structure of classification systems appropriate to the social sciences (Foskett, 1974), the natural sciences (Vickery, 1964), and the humanities (Langridge, 1976). Though Ranganathan had argued that a universal system using the faceted method allowed for flexibility in adapting to local contexts, the CRG went a step further by determining the facets themselves based on the nature of work in that field. For example, in the course of introducing the classification system for the social sciences, Foskett (1974) noted that the accumulation of knowledge in the social sciences occurred differently than in the sciences, so that assumptions about retrieval and the stability of terminology that applied in one domain could not be applied to the other (p. 26). However, Foskett (1974) argued that the classification designer did not need to be a greater expert in the field than its participants; rather, the position of the classification designer as a non-specialist within a field of many sub-specialties provided opportunities to create systems that facilitate interdisciplinary communication (p. 35). That is, the classification designer played an important role in making related knowledge domains intelligible to each other.

More recently, Hjørland and Albrechtsen (Hjørland, 2002; Hjørland & Albrechtsen, 1995) presented domain analysis as a new paradigm for information science, one which views users not as a collection of individuals but as a coherent domain of researchers and authors who participate in a particular discourse. By focusing on the domain as the unit of analysis, Hjørland (2004) hoped to address the "sadly ignored and neglected" state of subject knowledge by library and information science (p. 59) and described all other approaches as "attempts to avoid the necessary subject knowledge" involved in creating classifications (2013, pp. 178-179). The goal of domain analysis was to produce functional classification systems that served the needs of users by representing a knowledge domain through the theories and paradigms of that domain. For example, Hjørland (1998) took a historical approach to understanding the domain of psychology, finding that the multiple and changing epistemological foundations of the domain suggested many ways of organizing collections of psychology works. Hjørland and Albrechtsen (1995) argued that it was the classification designer's responsibility to understand the field well enough to recognize and represent the multiple, potentially conflicting schools of thought within the domain such that both the dominant and alternative paradigms of the domain were compatible with retrieval. However, critics of this approach have noted that domain analysis tends to represent this process as scientific discovery rather than an interpretive act of defining and organizing (Feinberg, 2007). That is, though domain analysis acknowledged that there are many ways to organize a domain of knowledge, the assumption remained that one organization of its divisions was most correct.

Similar to domain analysis, cognitive work analysis was an approach to information gathering focused on actual work situations, with the aim to produce systems that support information retrieval and sharing within the given work environment (Albrechtsen & Pejtersen, 2003). Designers used this approach in various settings to determine the nature of information needs and information flows in the course of work (Albrechtsen, Pejtersen, & Cleal, 2002; Huvila, 2006; Park, 2008b). In describing cognitive work analysis, Albrechtsen and Pejtersen (2003) explicitly compared the approach both to domain analysis and to the dominant model of creating systems through the application of guidelines and standards, stating that it shared the dynamic view of semantic structures with the former, and diverged from the latter's focus on universal principles (pp. 213-214). Cognitive work analysis was also a novel methodological framework in that its creators acknowledged the reciprocal relationship between the domain and the system; Albrechtsen and Pejtersen (2003) included within those systems that determined the articulation of work the classification systems built through such a process (p. 215). Therefore, the system itself had to be considered as one element of the work practices to be supported with ongoing system design.

Classification as Teamwork

Many of the canonical texts in classification theory have assumed a single classification designer with autonomy over design decisions. For example, Ranganathan's (1962) construction manual was addressed to an imagined singular librarian and discussed the work of the singular classification designer or classifier and Bliss (1929)

discussed the classificationist, singular, and explained his own approach to ordering the sciences as his individual project. Where other people appeared in these accounts was through their publications included in the system and the application of literary warrant, or through brief contacts to inquire on specific domain knowledge, such as the order of subjects in medical practice (Ranganthan, 1962, p. 41). While portrayals of the classification designer as a solitary worker emphasized the autonomy of classification design, accounts characterizing the work as collaborative and team-oriented have highlighted the multiple goals and socially embedded nature of these tasks.

The exceptions to the single, autonomous characterization of classification design have been the few scholars who place teamwork at the core of classification design. That these scholars have been few in number makes clear there is no trend towards an acknowledgement of team-based design overall; to wit, several recent scholars have used the image of an individual classification designer to emphasize the notion of authorial voice (Feinberg, 2011), to provide instruction to a singular reader of a text (Aitchison, Gilchrist, & Bawden, 2000), or, in historical treatments, to tie the values of a classification system to its named author (Olson, 2004). However, the representation of classification design as collaborative has been important to public institutions seeking to stress transparency and accountability (Young & Mandelstam, 2013) and in online projects in which classification design was highly distributed (Thornton & McDonald, 2012). Where classification design has been described as a cooperative task, the team has had two very different structures: either the classification designer was an information professional who supervised and directed the work of a group of workers with different

specializations, or the classification designer was one worker within a team of experts and users. Soergel's (1974) classification design manual exemplified the former position, the classification designer-as-manager. The latter position, classification designer-as-team member, was particularly important for the participatory design or mutual design movement in information system design (Albrechtsen & Jacob, 1998). In the next two paragraphs, I explain these two positions for classification designers within a team.

Soergel's (1974) classification designer delegated the implementation of his decisions to his staff and delegated the decisions regarding domain terms to the subject experts. The result was that the classification designer had two kinds of work: the people and project management skills of carrying out Soergel's many flow charts and the abstract design work that was left over once all the clerical and domain work was delegated out. Soergel presented his method for classification design in a series of detailed flow charts showing the elements of the design process. In the detailed version of "The Flow of Work" chart (p. 328ff), Soergel linked each step of the design process to the workers who were responsible for that aspect of the work—clerical staff for entering terms into the system, subject experts for identifying semantic links between terms, the classification designer for identifying appropriate sources of terminology. Soergel addressed the manual to the classification designer whose duties included defining the broad subject fields of the system, determining the structure of the system, and coordinating the work of the team. In particular, there were detailed instructions for the classification designer on how to extract information from subject experts. Specifically,

Soergel outlined a focus group approach that used successive groups of subject experts mediated by the information scientist to come to consensus on every term issue (p. 335).

A much different version of classification-as-teamwork appeared in accounts of participatory design approaches to the construction of classification systems. Albrechtsen and Jacob (1998) presented as a case study the work of the Ballerup Public Library to design a multimedia catalogue that represented the multiple points of view in their community. The team that created this project consisted of librarians, including reference and cataloging experts, and invited user groups to collaborate on various stages of the process, including identifying materials and evaluating the interface design. As a member of this team, the classification designer was "a participant in the process of knowledge production and mediation" and an "epistemic engineer" who built a system that facilitated interaction using multiple conceptual schemes (p. 296). By placing the classification designer in a flat team structure, this approach equated the classification designer's authority on final decisions to that of other team members, and the designer's contribution to the project also took on different meaning. The classification designer was not only someone who translated the group's design ideas into an organizing system, but was also someone the community recognized as shaping the library communication environment. This approach gave—or recognized—the classification designer as more active and visible than has been typical in rule-based approaches.

Classification as a Personal, Creative Act

The most recent work in classification research has also been the work that most stresses the agency and responsibility of the classification designer, particularly in contrast to rule-based approaches. While the CRG, the domain analysis approach, and the cognitive work analysis framework have moved the field from universal systems to specialized systems based on local knowledge, scholars have criticized these movements as merely reproducing the discourse of the discovery of truth common to early classification designers (Feinberg, 2007). Local approaches have acknowledged the dynamic nature of fields of knowledge, the limitations of scientific conclusions, and the plurality of paradigms within a domain (Hjørland & Albrechtsen, 1998) while presenting the classification as representing rather than participating in these phenomena. In contrast, recent scholars have argued that the classification designers' work is also dynamic and contradictory (Buckland, 2012; Feinberg, 2007; Keshet, 2011).

Scholars who have focused on the personal, social, and creative aspects of classification work have advocated transparency regarding the subjective nature of classification work (Feinberg, 2007; Mai, 2010) and suggested new possibilities for classification system design grounded in the idea of the classification designer as an active participant in the construction of knowledge (Albrechtsen & Jacob, 1998; Feinberg, 2011b; Olson, 1998; Theimer, 2012). Therefore, focusing on the classification designer as a socially and culturally embedded worker produced two complementary commentaries on the nature of classification work: the presence of bias and the possibility of creativity.

Classification research has a long tradition of examining systems with an aim to uncover their built-in harmful biases, though such studies typically focus on universal systems such as the Library of Congress Subject Headings and the Dewey Decimal Classification (Berman, 1971; Bowker & Star, 1999; Olson & Schlegl, 2001). Researchers read systems as the product of their particular social and cultural origins whose discriminatory status was more easily recognized at a distance (Olson, 1998). The outcome of this tradition of criticism has been a focus on contemporary systems and how they might be constructed to be more inclusive and less harmful (Albrechtsen & Jacob, 1998; Fox & Reece, 2012).

While Feinberg (2007) argued that even specialized classifications could not avoid the bias we now expect in universal systems, Mai (2010) took this limitation as a starting point. Assuming that there is no one right classification, Mai explored how the classification designer might establish their work among possible competing interpretations of a domain of knowledge. Mai argued for an explanation-based approach, in which the classification designer explicitly articulates the "system's premise, point of view, history, and objectives" (p. 634), rather than allow the system to stand as objective and ahistorical. Unlike those scholars who took the subjective nature of classification work as a challenge to produce the least flawed system (Szostak, 2008), Mai proposed a "principle of transparency" (2010, pp. 638-39) based on honesty and clarity between classification designers and users regarding the philosophical, pragmatic, and personal elements that shape these systems. Part of the classification designer's work, in this approach, was publicly documenting his or her own process and point of view.

Accepting the effect classification designers have on the classification system was not only a matter of recognizing and alleviating the negative effects; it also suggested opportunities for the classification designer to shape systems according to their particular insight and point of view. As the products of interpretive work, classification systems had the potential to communicate particular readings of a field of knowledge or a collection of items (Feinberg, 2011). Classification systems using this approach tended to be for web-based, standalone collections rather than universal collections. Classification designers could use this communicative potential to intervene where universal systems have failed, using creative means to undermine the Western, patriarchal bias embedded in existing systems (Olson, 1998) or to extend existing systems to cover marginalized topics such as world music (Nero, 2006).

Depending on the autonomy of the classification designer, this approach would appear more or less feasible. The classification designer-as-author would have the freedom to create a classification system that communicates a particular stance (Feinberg, 2007); a classification designer or cataloguer working within a network of interoperable systems would more likely be subject to pressures to produce consistent systems under deadlines (Hoffman, 2009). Regardless of the specific work environment of the classification designer, some scholars have argued that the current practice of information organization attracts individuals who are, or socializes individuals to be, isolationist, lacking in imagination, and overly serious (Theimer, 2012, p. 897). Furthermore, Fox and Reece (2012) noted that any information organization approach that allowed for exceptions and prioritized the accommodation of individual needs required a great deal of

deliberation and effort on the part of the classification designer. Giving classification designers the agency to produce systems that facilitate multiple types of access while respecting diverse individuals may have been an admirable goal, but Theimer (2012) described this same agency as a time consuming and weighty responsibility.

Summary

In the oldest and most taken for granted perspective, classification design is a process of applying principles and standards to a series of concepts. For the contemporary reader, this perspective appears as an extreme pole in the field of classification research, particularly in the form early classification research authors such as Bliss and Ranganathan expressed it. However, classification as the application of rules through a set process remains a common paradigm in the textbooks and manuals of classification work (e.g., Aitchison, Gilchrist, & Bawden, 2000). This perspective is particularly relevant to the ongoing development of automated classification construction techniques. Though it is an extreme, classification-as-rule-following remains a compelling version of how the classification designer carries out their work.

Criticisms against the classification-as-rule-following perspective have come in the form of practical as well as theoretical arguments. The idea that classification principles can or should be applied uniformly regardless of subject matter was the point of departure for domain specific classification design approaches. Domain analysis brought subject expertise to the center of the classification designer's skill set, demanding different interpretations of classification principles depending on the structure of a

domain of knowledge (Hjørland, 1998; Hjørland & Albrechtsen, 1995). Other traditions in classification research extended this idea of local exceptions further, studying the domain of knowledge within a single work environment in order to construct the most useful organizing system, as in cognitive work analysis (Albrechtsen & Pejtersen, 2003; Fidel & Pejtersen, 2004). Domain analysis and cognitive work analysis shared in common the perspective that universal classification principles were not sufficient to create usable classification systems given the variety of knowledge organization models and work tasks that exist in the world. The analytic steps that comprised these methods prioritized the accurate representation of a specific domain's model of their knowledge to the pragmatic end of supporting information sharing and retrieval among its members.

The most extreme objection against the classification-as-rule-following perspective has been that rule following is an illusory model that obscures the personal and social decisions made in the construction of classification systems. Scholars have pointed out that the classification principles provided cannot account for all the decisions made by a classification designer (Park, 2008b; Wild et al., 2009), that the design of the rules themselves were the product of a particular cultural worldview (Olson, 2004), and that appeals to the completeness of classification principles were too often used to justify harmful practices and systems (Mai, 2010; Olson, 1998).

The trend in the representation and prescription of classification design has been from universal, top-down procedures to local practices based on social and personal factors. While classification-as-rule-following positioned the classification designer as an expert in the processes and principles of the field, this perspective did not grant the

classification designer as much agency or responsibility as seen in more recent perspectives. The CRG, the domain analysis approach, and the cognitive work analysis framework shifted the focus from universal principles to the information needs and knowledge structures specific to a given field, domain, or work environment. In these perspectives, the classification designer applied subject expertise to produce systems individually suited to the collections and retrieval needs of these communities. In team-based and personal representations of classification design, the classification designer was not only responsible for understanding and representing a field of knowledge, but was accountable to others as the author of that system. In the move away from the rule-following approach, more recent theories have increasingly emphasized the situated, social, and personal elements of classification work.

Though I have presented these representations of classification work in a roughly chronological order, each remains an active position in contemporary classification research. Though some approaches are inherently contradictory—such as universalist, rule-based approaches and domain analysis—they continue to coexist across classification research journals and textbooks. The nature of the field of classification research is such that no one solution is likely to prove its superiority; the typical practice of publishing a classification system based on the theory as an appendix to a conceptual paper has invited evaluations on the basis of consistency and feasibility, not measurable outputs such as effectiveness in retrieval. The current situation of multiple, competing descriptive and prescriptive accounts of classification design requires more exploratory work regarding the actual practices and responsibilities of classification designers. The

general trend towards theories that emphasize the personal and local work of the classification designer suggests that such studies should not seek to generalize but rather should explore the interactions between the principles of classification theory and the social and cultural position of the classification designer.

SUMMARY: ON CLASSIFICATION DESIGN

The field of classification research has developed alongside changes in information technology and our understanding of the nature of language and meaning. New possibilities for the organization and retrieval of documents have suggested new methods of defining relationships among terms and between terms and documents. However, not all changes in the field have been progress towards better systems; awareness of the limitations and impacts of classification systems has been increasingly used to critique existing systems and theories of organization. Greater attention to the culturally and socially situated nature of classification design has motivated scholars to examine the biases of established universal systems and to build principles of transparency and creativity into new systems.

In this review, I described how the field of classification research depicts the classification design: the decisions made, the information needed, and the skills applied. While the majority of scholarship in this field speaks in generalities and keeps the minutiae of classification design itself at arm's length, trends and points of contradiction are evident. In each of the concepts I compared, the field shows major divides in priorities and conclusions: whether the core of classification is structure or meaning,

whether a classification designer can or should rely solely on the literature to determine terminology, and whether it is global principles or the individual classification designer's knowledge and insight that create a good system. These debates are not necessarily specific to classification research. The tension between form and content, for example, is a theme throughout all domains that guide and critique the creation of artifacts. The primacy of rules or an individual's unique perspective is particularly relevant to design studies; while the former inaccurately predicts uniform outputs of all designers, the latter fails to describe how work actually gets done (Schön, 1988).

Each of the competing opinions remains a relevant position in contemporary classification theory. Proving the merit of any approach over another would require not only publishing a possible classification schema along with the conceptual argument (as in Hjørland, 1998; Wang et al., 2008), but also implementing these schemas in live, competing systems. However, determining the performance of any system in practice is likely to be as contentious as determining the validity of any abstract approach; classification scholars do not agree on whether retrieval metrics such as precision and recall, or long-term outcomes such as changing users' mental models of a knowledge domain, are theoretically sound goals. To continue the comparison to design studies, classification research has the challenge of comparing approaches by means of blueprints, without the opportunity to construct the buildings themselves. Even if classification designers wished to recreate a "building" in a new way, the work is largely intellectual and the design explicit; recreating the Dewey Decimal System would merely be a matter

of copying the schedule. It is unlikely that classification designers would ever create truly competing systems for such comparisons to occur.

The intent of this review has been to focus on what is known about how classification designers go about creating their systems. However, the review also reveals the limitations of the dominant narratives of classification practice in this field. The field of classification research offers no clear consensus on how classification design ideally proceeds, let alone how it occurs in practice. What the field does offer is multiple perspectives on which elements—decisions, information, and skills—are most important to this process. The strength of contemporary scholarship in classification is the link between ways of approaching design, such as the types of information consulted, and principles regarding the purpose and potential of classification systems. The current approaches do not reveal how classification designers negotiate conflicts among these principles and approaches in daily work. For example, if literary warrant is insufficient to make semantic decisions (Fidel, 1994; Mai, 2005), how do classification designers justify and reconcile—to themselves, their institutions, and their users—the inevitable use of personal and contextual information with the belief that the system should represent the documents without bias? Similarly, at what point (if any) in a domain analysis process do classification designers encounter the impact of their systems on the domain being organized, and how do they seek to minimize or justify this distortion?

At this time, classification research does not have the answers to these questions. The literature I review above largely focuses on how classification design should proceed (e.g., Bliss, 1929; Hjørland, 2002; Ranganathan, 1959) or presents retrospective accounts

of model processes (e.g., Albrechtsen & Pejtersen, 2003; Nero, 2006). The literature does not describe how classification systems come to take on the values or bias that draw the concern of classification scholars (Berman, 1971; Bowker & Star, 1999; Olson, 2001b). To close this review, I shift to a complementary field of research—infrastructure studies—that addresses similar questions of classification systems and other infrastructures. Whereas classification research has largely separated the study of system design from system consequences, in infrastructure studies this connection is a major focus of analysis. By reviewing how infrastructure studies differs from classification research, I present a complementary research lens with the potential to uncover the missing connection between design and consequences.

The Infrastructure Studies Lens

Infrastructure studies examines systems such as telecommunication networks, highways, and information infrastructures—including classification system (Bowker & Star, 1999; Millerand & Bowker, 2007). Noting that information infrastructures such as classification systems are invisible—fading beyond our notice during our interactions with the collections and actions they organize—scholars have sought to uncover their values and consequences (Bowker, Baker, Millerand, & Ribes, 2010; Bowker & Star, 1999; Edwards, Jackson, Bowker, & Knobel, 2007; Star, 1999; Star & Ruhleder, 1996). For example, Millerand and Bowker's (2007) study of scientific metadata standards revealed how adoption of a single metadata standard generated controversy among different sets of actors in the scientific community—standard developers and information

managers. The metadata standard itself would not reveal these controversies, nor would a retroactive account from the perspective of metadata standard development. Instead, Millerand and Bowker took an ethnographic approach to detail how the process of enacting the standard intersected with existing organizational structures and scientific practices. This approach ultimately provided insight into how the metadata standard came to be as it was, how it supported some existing organizational structures, and how it required changes to other organizational structures. Over time, such standards might seem inevitable and their consequences natural; only by close examination have scholars linked consequences to the choices of designers and the context of design.

Primarily, infrastructure studies uncovers values and consequences through the methodological approach of "infrastructural inversion," (Star & Ruhleder, 1996, p. 113) which calls attention to the practices that support an infrastructure, rather than the activities the infrastructure makes possible. For example, in the case of Millerand and Bowker (2007), the focus was on the development and enactment of the metadata standard, not how the standard supported scientific collaboration. A focus on practices reveals elements of the infrastructure that we miss by examining infrastructure as an object (Star & Ruhleder, 1996). To view classification design as one such practice would mean shifting from thinking about classification systems as supporting organization and retrieval to thinking of classification design as a practice that "hold[s] together the functioning" of the classification system (Ribes & Lee, 2010, p. 238). This sometimes requires "spelunking" into black-boxed elements of design such as code and system measurements (Ribes, 2014, p. 162). In Ribes (2014), a close look at the computational

infrastructure for the Large Hadron Collider revealed the hidden work—deploying, rejecting, and combining different tools and metrics—required to assess a system's ability to scale to support cutting edge particle physics research. While the Large Hadron Collider and the research it supports typically takes center stage, infrastructural inversion revealed the multitude of invisible practices and actors that determined how its computations were processed.

Three elements of infrastructural inversion in particular offer promising complements to the dominant modes of studying classification systems: a focus on how systems change (in contrast to studying ways that systems change the world around them); a focus on maintenance work (in contrast to studying initial design); and a focus on how systems are embedded within and interdependent with other systems (in contrast to studying a system in isolation) (Star & Ruhleder, 1996). In the following paragraphs, I examine each of these elements in turn with reference to their divergence from the classification research literature.

With regards to change, scholars have examined how systems facilitate or resist changes in the world around them, such as medical practices and racial apartheid (Bowker & Star, 1999). Looking within classification systems, scholars have taken the inevitability of classification system change as a point of concern, such as by noting the need for systems to adapt to contemporary contexts (Olson, 1998) and by identifying the challenges to address the phenomenon of obsolescence (Buckland, 2012). However, scholars from Bliss (1929) to Hjørland (2002) typically discuss classification design as system creation. When scholars discuss system change, they typically do so from the

perspective of outsiders extending an established system (e.g., Nero, 2006; Olson, 1998) and rarely as insiders revising an established system (e.g., Young & Mandelstam, 2013). The result is that change in classification systems appears as a provocative or extraordinary event, while infrastructure studies centers change as important to understanding the relationship of a system to the practices and actors around it. Focusing on system creation alone leaves out the decisions classification designers make to adjust a system to unforeseen user needs and changes in intersecting infrastructures, among other factors.

Similarly, the focus in classification research literature on system creation masks the importance and ubiquity of system maintenance. A notable exception is Soergel's (1974) manual which includes "maintenance" in the title and devotes a chapter to updating and maintaining terminology. As with a focus on creation without change, a focus on creation without maintenance leaves out types of decisions relevant to the character of the classification system. For example, without an account of regular maintenance, we would have no indication of how classification designers intend to resolve shifts in terms' meaning over time, a critical concern for long-lived systems (Buckland, 2012). How do classification designers set a threshold of change in a term's meaning or recognize that threshold? How do classification designers prioritize changing a term's meaning against the impact of misrepresenting or making invisible older items in the collection? Furthermore, without a focus on maintenance, we would not notice designers' choices that keep the system from changing noticeably, such as maintenance

choices that anticipate trends in advance, adding terms before users experience a failed search for a popular referent.

Finally, scholars typically discuss classification systems as isolated systems. This isolated view has been especially common among conceptual accounts that focus on ideal design (e.g., Bliss, 1929; Mills, 2004; Ranganathan, 1961). Notable exceptions include scholars who presented design approaches such as cognitive work analysis (Albrechtsen & Pejtersen, 2003) that focused on the specifics of the environment in which the system is deployed. As with the lack of study of change and maintenance, the lack of study of embeddedness leaves out types of decisions classification designers are likely to make or factors that will influence those decisions. For example, a classification designer managing medical subject terms would likely consider interoperability with established systems such as the widely-used Medical Subject Headings vocabulary (Lipscomb, 2000). Alternatively, a classification designer wishing to integrate terms from other languages might have to compromise based on which characters the database will accept. Without a focus on embeddedness, we do not know how these interactions impact the character of a classification system or how classification designers decide how to manage these interactions.

These three elements from infrastructural inversion are not adequately explained in classification research literature. Studying classification systems as infrastructure expands the focus of study from the system itself to its relationships to practices and other infrastructures and makes it more likely that we will account for the classification design decisions that create the character of the classification system. However, the potential of

this approach must be combined with classification theory itself. As Ribes (2004) noted in his description of the infrastructural inversion of spelunking into computer code, there are choices made in system design that are visible only with expertise in that mode of design. In other words, to study the decisions classification designers make, it is necessary to recognize those decisions. Before turning to my research questions, I briefly summarize the types of questions classification research literature suggests are unresolved.

Classification research literature presents a host of different approaches and assumptions regarding the potential and purpose of classification systems; what remains unclear is how these approaches are embodied in the daily decisions of classification designers. For example, if designers see the purpose of classification system design as producing systems that users can easily use to sort and retrieve documents, do these designers then apply user warrant according to established rules and without judgment? In such a case, the character of the resulting system is purely the result of the corpus of user terminology that the designers apply as user warrant (e.g., racist users produce a racist system). Scholars have stated that such a reality is unlikely, even in very controlled classification design settings (Mai, 2005; Theimer, 2012). Furthermore, this assumption abandons hope that classification designers can address harmful bias in classification systems and accepts the inevitability that classification design should be performed by computers. However, if it is the case that designers are not merely conduits of terminology and rules, but are thoughtful actors aware that classification systems can embody potentially harmful biases (Bowker & Star, 1999; Feinberg, 2007; Mai, 2010),

how do they make their choices? Do they, perhaps, make creative, intentional choices (Feinberg, 2011a; Theimer, 2012)? In this case, the character of the resulting system is the result of a larger set of factors including the corpus of terminology, the designers' sense of the purpose of the system, and the impact of intersecting infrastructures. To understand this more likely but less examined version of classification design, I turn to the study of classification design as it actually occurs.

Research Questions

The following study addresses classification research's lack of connection between classification methods and classification consequences by focusing on classification designers' decisions. Specifically, I adopt the approach of infrastructural inversion and examine classification designers' decisions to reveal how a classification system changes, is subject to maintenance, and is embedded in other infrastructures. By examining daily, ongoing design decisions, I can account for change and maintenance. By examining a live system, actively deployed and in use, I can account for its embeddedness in other infrastructures. Therefore, my research questions focus on how designers make daily decisions in a live system. Furthermore, I examine how designers understand their decisions in the context of the classification system, including how their decisions instantiate or violate different views of the purpose of the system and whether they consider factors beyond the established choices among types of warrant.   In answering these questions, I contribute to the pressing concern in classification theory we

do not understand how classification design produces classification systems embodying particular values and biases.

In the next section, Methods, I detail how I sought to answer the following research questions that address the gap between classification methods and classification consequences:

1. What factors do classification designers consider in making daily decisions on terms and term relationships?

2. How do classification designers negotiate multiple or conflicting purposes for the classification system?

3. To what extent do rules determine the character of large scale classification systems?

Methods

I chose a research site and applied ethnographic methods to answer my research questions. In this chapter, I describe the site selection, my data collection methods, and my data analysis.

CHOICE OF RESEARCH SITE

I chose a site that served as an example of a successful large-scale user-driven classification design project. The most salient characteristics to its suitability as an example were its stability over years of use, the functionality in regards to browsing and searching for works in the resulting system, and the continuing interest of volunteers in contributing to the system. Its stability over years of use indicated that the system, as I studied it, was representative of a sustainable classification design approach; it was not a short-lived experiment or provocation—indeed, at this time, it continues to operate as I describe it. Its functionality in regards to browsing and searching was comparable to large-scale systems from library databases to shopping interfaces. The continuing interest of volunteers in contributing to the system indicated that this community saw the design work as appropriate to the functionality it offered. I found this assumption to be true in my discussions with its designers, who viewed the work as demanding and at times complex, but integral to the use of the collection.

The Fanwork Repository (TFR) was a non-profit that was funded and built—in terms of code, contents, and policy—by its users. It underwent constant maintenance and development from volunteers who debugged the site and added features, chiefly to enable the expansion of its user base and scope. One such team of volunteers continually designed the classification system. At the time of writing, TFR included over two million fanworks (fan-created documents such as writing, audio, and artwork) and had a monthly

traffic of seven million visitors. Users added tens of thousands of new works every month, each with a handful of tags attached. Users created tags—descriptive terms attached to works as means of identification and categorization—with very few limitations. Specifically, the users were free to enter tags containing text, numbers, and many special characters (e.g., "&") up to 100 characters in length.

The classification system for this site—described in more detail in the section Classification Design at The Fanfiction Repository—required constant maintenance by a team of trained volunteers, known as "wranglers." In the following four subsections, I describe in more detail the characteristics of this site. The first three characteristics—that it is a live system, that it is a growing, and current system, and that it is a system with familiar concerns—speak to the generalizability of my findings to other classification systems. The last—that it is a system of explicit reflection—speaks to the advantage in researching this site over one that is more familiar to classification research. After detailing these four characteristics, I note the limitations of directing my research questions towards a single system.

A live system

As I note in my literature review, one of the major impediments to productive debate across schools of thought in classification theory is that scholars discuss design as an ideal concept (in the description of a championed approach) or retroactively in simplified, rationalized, and uncontroversial accounts of past work. Furthermore, in each of these cases, the design in question is typically only the initial creation of the system, with the ongoing work of maintenance, growth, and revision absent from debate. Similar pragmatic limitations on research approach are common in infrastructure studies, in which scholars examine a system in use (Bowker & Star, 1999) or in design (Ribes &

Bowker, 2009) but seldom both. To capture a more complete view of classification design, I chose a live system—neither a theoretical, ideal system being designed for the purposes of demonstration nor a planned system not yet deployed to users. In the design decisions around such a system, I was more likely to find interactions and compromises among user expectations; the collection itself; interdependent technologies and their constraints; and the classification designers and their inherited principles and assumptions.

A growing, current system

Criticisms of classification systems based on their dissonance with contemporary values, such as gender equality, are usually presented in the form of close readings and analysis of the written system (Berman, 1971; Bowker & Star, 1999; Olson & Schlegl, 2008). However, live classification systems—especially those for ongoing and growing collections—are not finished texts, but are instead subject to ongoing revision, and designers return to engage with these issues of values and changing context (Olson, 1998; Young & Mandelstam, 2013). The long lifespan of infrastructure necessitates and provides opportunities for "design-after-design" by a variety of actors as values and context change over time, so that the use and character of a system undergoes periodic change (Bjögvinsson, Ehn, & Hillgren, 2012). Such processes of maintenance are integral to the functioning and character of infrastructural systems, but are often overlooked by developers and users (Bowker et al., 2010; Ribes & Finholt, 2009; Star & Ruhleder, 1996). As a live system that is both growing and current, TFR provides opportunities to examine the work that designers do to adapt to current user values and context. Capturing this aspect of classification design is particularly important for the question of how the design process contributes to the values or bias of the live system. The system and its

ongoing revision work do not represent a single cultural moment; contemporary designers are familiar with the same concerns as those scholars pose in classification theory, and maintenance work means reflecting on both original intention and current concerns. Revision work in particular addresses ongoing, sticky questions in classification theory, such as the inevitability of obsolescence and its technical challenges (Buckland, 2012).

A system with familiar concerns

The system at TFR is similar in many ways to those systems held up as exemplars in classification research: library classification systems such as the Library of Congress Subject Headings and the Dewey Decimal Classification (Adler & Tennis, 2013; Buckland, 1999; Olson, 1998; Svenonius, 2000). These exemplars provide a common ground within classification research and to related communities such as infrastructure studies. As a result systems—classification systems for the organization and retrieval of written work—are continually invoked in classification research and infrastructure studies as they are familiar objects through the subject of scholarship and through our shared everyday interactions (Millerand & Bowker, 2007; Ribes & Bowker, 2009; Star & Ruhleder, 1996). My research site shares several core aspects with such systems, providing opportunities to add to the research on these systems and to capitalize on our shared familiarity with them.

For example, TFR enacts classification design relationships such as broader, narrower, and related terms—an aspect it shares with classification systems from library classification systems to scientific ontologies. Moreover, the express aim of TFR's design is retrieval, like the above systems, and unlike systems built as provocations to inspire new ways of thinking about classification (Feinberg, Carter, & Bullard, 2014a, 2014b). In this respect, it aligns to the example of library classification in that it ultimately organizes

a set of objects or documents (not merely a set of concepts without definite manifestations); TFR primarily organizes a set of written documents, as audio or image collections carry a distinct set of challenges and approaches. Finally, like library classification systems, TFR covers a broad range of subjects, includes many kinds of documents, and serves a diverse population of potential users. As a result, TFR's classification designers cannot reasonably check each decision they make against the universe of possible effects and stakeholders, necessitating the development of general guidelines and principles. In sum, TFR's concerns are familiar to classification scholars.

A system of explicit reflection and discussion

While the previous characteristics relate to the substance of my research questions and relevance to established theory, this last characteristic addresses the pragmatics of answering these questions in the scope of a dissertation and without unduly disrupting the design work I wish to observe. My site is one in which uncovering the processes and reasoning behind classification design decisions would not change how classification designers do their work. Commonly, a single worker or small group of workers in a larger organization performs classification design and maintenance (Park, 2008a; Wild et al., 2009; Young & Mandelstam, 2013). For small or slowly growing systems, a consult performs the design and periodically returns to the system, or a worker with a broader or related job title (such as "librarian") performs the design as a part-time task. When the work is often solitary and infrequent, it can be difficult to get designers to articulate their reasoning and may be disruptive to the work process to elicit the reflection necessary to tease out these aspects. It is not surprising, then, that many of the accounts of classification design rely on recall of decisions and are largely rationalized versions of the

process. That is, authors' retrospective accounts are likely to represent what made sense and lose some of the trade-offs and messy decisions made along the way.

In this respect, TFR as a design site has two advantages that make it an ideal site for capturing the process behind decisions without radically disrupting the design process. First, wranglers typically explained or documented their decisions for others. Hundreds of wranglers performed the design work at this site. The inevitable churn or change-over in the makeup of this volunteer group meant that wranglers constantly explained and interrogated the guidelines. Because wranglers lacked a common professional or educational background, the training process identified explicitly many elements of classification processes that might go unsaid in other settings. Wranglers at TFR were accustomed to explaining their decisions as part of the learning process for new volunteers.

Second, as a large and distributed design site, TFR had its own processes of self-understanding and monitoring, so that the project's own coordination practices can double as sources of data or means of data collection (Ribes, 2014). Specifically, the team's large size and changing membership necessitated routine documentation; as a result, workers were accustomed to recording their decisions to notify other workers, leaving traces of discussion and compromise I used to help make sense of their decisions. Similarly, coordination practices such as training exercises (Bullard, 2016) and internal meeting notes consolidated the history and concerns of the community for their own uses in continuing the work, and provided further material for me to understand long-standing issues and consensus in the community. Furthermore, because control was dispersed across so many designers, the perspective of any one designer—given their domain of responsibility, the peculiarities of that section of the collection, and the attitudes of the domain's engaged audience—could potentially disrupt the consensus of the rest of the

group, forcing the community to routinely re-argue guidelines and principles from base assumptions. Specifically, these discussions occurred through structured and unstructured threads in the group's mailing list. Such threads document the variety of positions within the community and how they came to consensus on site-wide policies.

A single system

This section on site choice has illustrated how TFR serves as an example generalizable to other classification design domains. In addition to sharing key characteristics with other classification design domains, TFR also interacted with parallel classification systems. Notably, users built TFR in part to offer an alternative to commercially-oriented websites such as fanfiction.net (which used a top-down, limited controlled vocabulary) and websites devoted to smaller fan communities (which used local terms or uncontrolled tags). By studying TFR as a single system in isolation rather than within a network of related infrastructures, I risked missing how interactions among these systems affected the design of TFR.

In the next section, I provide more detail on the specifics of classification design at TFR. These specifics differentiate TFR from more familiar systems such as library classification and are necessary to understand the details of data collection that follow.

CLASSIFICATION DESIGN AT TFR

This section serves to detail the classification design process and the elements of the system I reference in my findings chapters. In this section I introduce the genre of knowledge organization instantiated at the TFR. After situating this genre of classification design within established genres of knowledge organization, I describe the particulars of TFR's approach. I first introduce the wrangling position. I then outline the

structure of tags at TFR and then present a narrative account the designers' process of interacting with those tags.

Curated Folksonomies

The tagging framework for TFR is a "curated folksonomy," "tag gardening," or "democratic indexing" a relatively new genre of knowledge organization that approximates the utility of a controlled vocabulary or thesaurus through the management and manipulation of a user-generated folksonomy (Hidderley & Rafferty, 1997; Peters & Weller, 2008). Folksonomies, traditionally, are aggregate systems of tags users assign to items, with few limitations (such as the use of special characters and punctuation). In retrieval, folksonomies return all objects tagged with the identical string and do not facilitate broader- and narrower-term relationships.

Folksonomies are ubiquitous across contemporary digital collections, but common wisdom and extensive research reveal at least four inherent shortcomings. First, folksonomies suffer from user error, so that misspellings divide otherwise identical tags in retrieval. Second, folksonomies suffer from diversity of user perspectives, so that small differences in syntax ("truck" instead of "trucks"), specificity ("Ford F150" and "truck"), and language ("color" instead of "colour") produce different retrieval sets, negatively impacting recall. Conversely, folksonomies suffer from convergence through homographs, so that "orange" (color) and "orange" (fruit) are conflated in retrieval, negatively impacting precision. Third, research has found that folksonomies overwhelmingly feature self-directed tags, such as "to read," producing aggregate sets that serve little function beyond an individual's collection (Sa Golder & Huberman, 2006; Munk & Mørk, 2007). Fourth, while folksonomies harness the energy and interest of users in managing their own objects to produce a system that extends the retrieval

capability beyond full-text searching in cases where more intentional or top-down organization is not feasible, the distribution of tags in folksonomies tends to follow a long-tail distribution in which some tags are very widely used while the majority of tags have few uses (Sa Golder & Huberman, 2006; Munk & Mørk, 2007). Therefore, the tags which do indicate consensus—those on the higher end of the long-tail distribution—are either so general as to be meaningless in retrieval or are merely repetitions of terms available otherwise in metadata, such as titles (Munk & Mørk, 2007). Interface design for folksonomies can help to alleviate some of these common shortcomings. For example, sites such as Delicious.com produce suggestions from the existing folksonomy, speeding convergence on terms through imitation, which addresses divergence in word choice (Sa Golder & Huberman, 2006; Munk & Mørk, 2007).

In contrast to classic or unregulated folksonomies, curated folksonomies take the aggregate tags users add as a starting point and use expert or collective decision making to identify and alleviate problems with synonyms and homographs. Scholars have introduced similar approaches as "tag gardening" (Peters & Weller, 2008) or "democratic indexing" (Hidderley & Rafferty, 1997; Rafferty & Hidderley, 2007). Regardless of the name, the basic tenets of these systems are as follows:

1. Users create tags.

2. Some intentional agent combines synonymous tags, improving recall.

3. Some intentional agent differentiates homographic tags, improving precision.

Curated folksonomies are primarily reactive: unlike terms in traditional, top-down classification design approaches, user activity drives most of the terms in a curated folksonomy, and intentional design follows user action. Curated folksonomies are particularly appropriate for user-generated collections that are quickly growing, in which currency is of particular concern, for which professional classification design is too

costly, and in which users are particularly motivated and suited to engage in organizing work.

In classification research—as tag gardening or democratic indexing—the curated folksonomy approach has been largely theoretical (Hidderley & Rafferty, 1997; Peters & Weller, 2008; Rafferty & Hidderley, 2007). At this time, two notable examples of large, user-generated collections instantiate versions of the curated folksonomy approach: LibraryThing and Stack Overflow. In the next section, I outline these two instantiations of the approach. The section following describes TFR as a third instantiation of this approach.

Curated Folksonomy at LibraryThing and Stack Overflow

LibraryThing, a user-generated database of books, uses a curated folksonomy to manage tags users apply to books, often indicating content ("animals"), genre ("satire"), or personal relevance ("summer reading"). In LibraryThing's system of "Tag combining" ("Tag combining," n.d.), all interested users can identify synonymous tags for combining, identify wrongly combined tags for separating, and vote in both types of decisions. The tag combining system is notably strict and follows a relatively narrow sense of synonymy, as described on LibraryThing's wiki:

> Tag combination is driven by a single basic rule: Tags should be combined only when they are the same in both meaning and usage on LibraryThing.
>
> Examples:
>
> - There is no discernible difference in either the use or meaning when it comes to terms like "wwii," "ww2" and "world war two."
>
> - While some might claim they are synonyms, tags like LGBT and GLBT have very different top books. It's likely they encode differences in perspective or identity. ("Tag combining," n.d.)

In LibraryThing, tag combining does not equate plurals to singular versions nor all abbreviations to full terms, in expectation that some cases actually indicate different meanings. Tags that users "combine" on LibraryThing redirect to the page for the preferred version of the term. This page displays the collected combined terms and all books tagged with any variant of the term. Usages of variant versions of the tag remain unchanged on book and user pages; combining decisions only affect site-wide retrieval. This design choice keeps intact users' own organizing schemas within their collections, particularly important for tag variants in other languages (e.g., "ciencia ficcion" and "science fiction").

The second notable example of an active curated folksonomy is Stack Overflow, a question and answer site for computer programming. All Stack Overflow questions include one or more tags. Tags are useful for retrieval of related questions and have information pages (similar to scope notes detailing the meaning and application of terms in traditional controlled vocabulary systems) that give background information. Users with higher algorithmic "reputation" on the site can edit tag information pages, can identify tags that are synonymous to "master" tags, and can vote for or against these connections. Tags users vote to be synonymous with "master" tags are "automatically and silently remapped" by the system ("Tag Folksonomy and Tag Synonyms," 2010) to their master tags; the system changes the synonymous forms of tags to the preferred form throughout the site.

These two examples of curated folksonomy share basic tenets—that users drive the creation of new tags and that some users make synonymous tags equivalent for retrieval—although the philosophy of the folksonomies differ. Specifically, at LibraryThing, tagging instructions present tagging as a self-directed, personal information management tool: "Once you have a hundred books or so, you need some

way to organize them" ("Some LibraryThing Concepts," n.d.). In this case, the folksonomy is truly "a function of the total sum of *persononmies*" (Munk & Mørk, 2007) (emphasis added), or individual tag sets users create for the organization and material of a personal collection. Particularly with the augmented functionality the curated folksonomy confers, the aggregate has the secondary function of supporting retrieval and discovery across the entire site. In contrast, Stack Overflow, as a question and answer site, is primarily outward-facing, and the community intends for tags to help potential answerers monitor questions in their areas of expertise. The curated folksonomy of Stack Overflow furthers questioners' efforts to have relevant answers see their tags by aligning variant tags with expert-preferred synonyms. These two systems illustrate different design decisions in the structure of a curated folksonomy appropriate to the different purposes of each site. In the following sections, I outline the structure of the curated folksonomy at TFR. This system is retrieval-oriented, like Stack Overflow, while preserving individual user choices, as in LibraryThing.

Curated Folksonomy at TFR

The form of curated folksonomy of TFR fell between the examples of LibraryThing and Stack Overflow. As with objects in LibraryThing, the tagged objects in TFR were primarily textual works and the curated folksonomy actions did not change user-chosen variants, only equated terms in retrieval. As with Stack Overflow, the primary activity of the site was outward-facing—authors to readers—and the nominal purpose of tagging and the curated folksonomy was to increase the visibility of user-generated content to relevant readers. Of the three classification systems, TFR was arguably the most selective with regards to which users participate in curated folksonomy design; whereas LibraryThing allows all users to nominate and vote, and Stack Overflow

allows established users to nominate and vote, TFR had a small (~200) team of users—referred to here on as "wranglers"—who completed a recruitment and training process before receiving access and permissions to the curated folksonomy interface. Wranglers did not nominate and vote on decisions within the curated folksonomy; rather, each wrangler was responsible for sections of the site and made changes to the folksonomy with relative autonomy.

The three following subsections introduce wranglers as actors in the system, detail the form of the curated folksonomy at TFR, and detail the process by which wranglers shape the system, In these sections, I describe the curated folksonomy as wranglers understood it. I use the wranglers' own terms throughout, capitalizing these emic terms as proper nouns from the system (e.g., Fandom, Unfilterable).

Wranglers at TFR

TFR users volunteered for wrangling positions. The organization recruited volunteers in recruitment drives twice a year or according to the volume of unattended or understaffed domains. In these periods, 30 to 100 volunteers joined (or re-joined) as wranglers. Once a year, organizers required volunteers to indicate whether they intended to continue volunteering in the following year and retired individuals as appropriate, though volunteers also reduced their commitments or left the organization throughout the year. Between 2012 and 2016, the number of classification volunteers varied between 200 and 240 individuals.

During recruitment drives, organizers posted calls for volunteers on the TFR site and requested that the message be spread by users across social media. Only during recruitment drives were volunteers eligible to apply to the organization. Calls for wranglers did not request that volunteers have experience with metadata management,

but often indicated which areas are in immediate need of wrangling to attract volunteers with interest in and familiarity with those source materials and their fandoms. The term "fandom" typically refers to a community of fans and, in this context, a fandom was the fan community around a particular work or set of works (e.g., *Romeo & Juliet* fandom, NFL fandom). Interested volunteers filled out an online application. The organization did not disclose how many volunteer applications—if any—were rejected. Designated senior volunteers, referred to as staff, notified new volunteers of their acceptance, assisted them in setting up accounts on various platforms used for distributed work, and instructed them to begin following the training materials within the organization's private wiki.

The training materials introduced new volunteers to wrangler terminology, organizational policies, the technical interface of the wrangling system, and the wrangling process, including the governing guidelines. Accepting a volunteer wrangler position required a minimum commitment of one hour per week; the participants in this study wrangled for an average of six hours (median four) per week.

Tag Structure at TFR

Tags at TFR were user-defined within a structure of four tag categories: Fandom, Character, Relationship, and Freeform. The tagging interface on posting works forced authors to indicate the category to which each of their tags belonged. Beyond that choice, users were only restricted by character-length limitations for individual tags and by limitations on certain special characters. For example, though users were encouraged to use the Character tag field to enter the characters present in their story, the interface did not prevent the user from entering "short story," "The Fellowship of the Ring," or "I wrote this at 4am" in this field. Fandom tags had a special role in organizing other tags in the wrangling interface in that Character, Relationship, and Freeform tags had a

"Fandom" attribute. The wrangling interface suggested one or more possible entries for the Fandom attribute derived from the Fandom(s) added to the same work, but this suggested Fandom association had no effect on retrieval; its only function was to organize tags within the wrangling interface and as a decision-making aid for wranglers. Users other than a work's author could also add tags as Bookmarks; users could search for fanworks by Bookmark tags, but this feature was not as prominent in display as the authors' tags. Users entered Bookmark tags into a single tag field. The tags were initially Unsorted and only took on one of the four categories when assigned to one by a wrangler.

Tags entered by authors or sorted by wranglers into a category existed in one of four states: Unwrangled, Unfilterable, Canonical, and Synonym. Users could interact with tags of any state both in display and retrieval. These states indicated to what extent the tag had been processed through the curated folksonomy and what—if any—connections it had to other tags in the system. Unwrangled tags belong to a category but do not have assigned Fandoms. Unfilterable tags had a Fandom (such as "Harry Potter – J.K. Rowling" or "Arthurian Mythology"). Tags could have multiple associated Fandoms. Canonical tags were preferred tags, in the terminology of controlled vocabularies, and were the most visible tags in the user's display. Canonical tags appeared as autocomplete suggestions when users were entering tags into a category, and canonical tags associated with a fandom already indicated on the new work were privileged in the order of autocomplete suggestions. Canonical tags could have an unlimited number of Synonym tags associated with them and return all works tagged with their Synonym tags. A Synonym tag redirected to a single Canonical tag; it still appeared on works as users entered it, but it did not appear in autocomplete. Additionally, when users selected a Synonym tag link, the Canonical tag was displayed instead. Tags which were neither Synonym nor Canonical could be "Unwrangleable," a designation that

removed the tag from default wrangling interfaces but did not prevent users from adding the tag to works or retrieving works through the tag. Broader- and narrower-term relationships only existed among Canonical tags: Canonical tags could have one more or Metatags (Canonical tags with broader meanings) and one or more Subtags (Canonical tags with narrower meanings). Figure 1 illustrates the process of tags moving from Unsorted to Synonym or Canonical. States are labelled with associated wrangler actions below. Tags often remained Unfilterable indefinitely, but Unsorted and Unwrangled were intended to be transitional states.
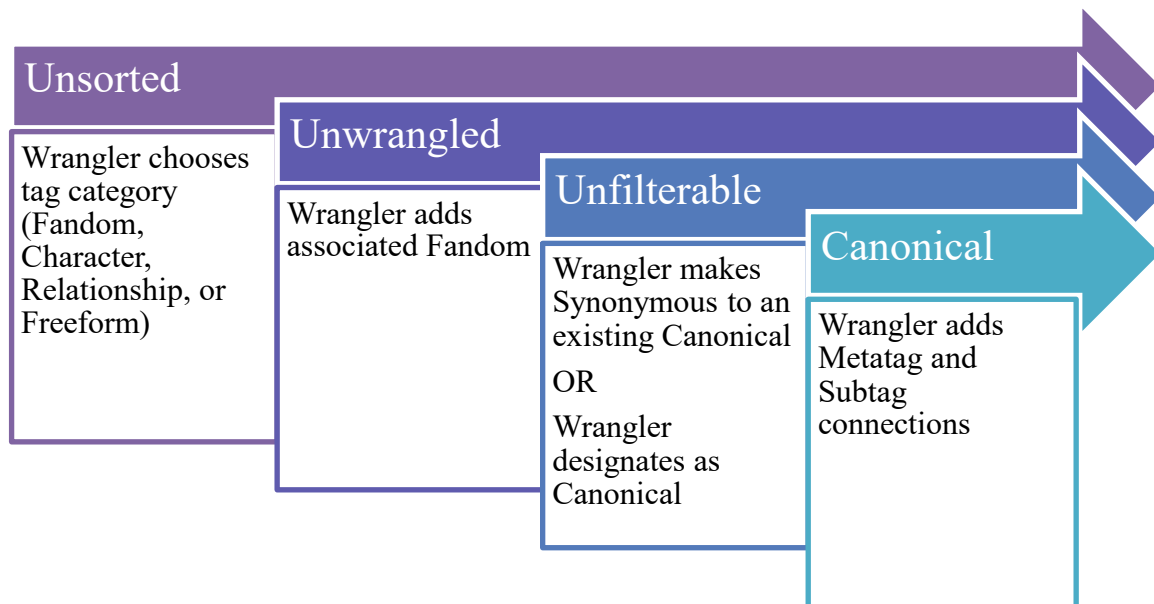
Figure 1 Wrangler interactions with tag states

As the wranglers process the tags, the tags pass through different tag states. Table 1 lists the properties of each tag state.

	Unsorted	Unwrangled	Unfilterable	Synonym	Canonical
Appears on Works	✓	✓	✓	✓	✓
Has a Tag Category		✓	✓	✓	✓
Has an Assigned Fandom			✓	✓	✓
Can Have Synonyms					✓
Can Have Meta- and Subtags					✓
Appears in Autocomplete					✓
Appears as Title on Search Pages					✓

Table 1 Tag state characteristics

Links to Metatags returned all works tagged with the Metatag Canonical, its Synonym tags, all Subtag Canonicals, and their Synonym tags. Subtags could have multiple Metatags. Figure 2 diagrams the interaction of Meta- and Subtags.
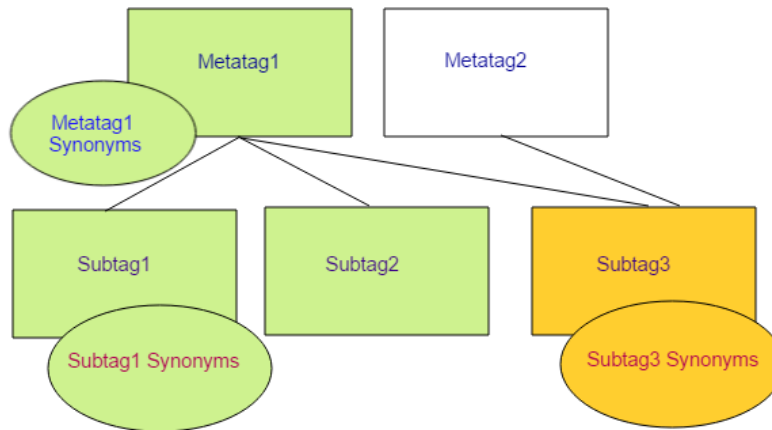
Figure 2 Meta- and Subtag interactions in retrieval

The green- and orange-highlighted tags indicate those tags that would be used to return works if Metatag1 or Metatag1 Synonyms was searched or clicked as a link. The orange-highlighted tags indicate those that would be used to return works if Subtag3 or Subtag3 Synonyms was searched or selected as a link.

The above example describes wrangling in terms of the function of tags and the curated folksonomy. The next section describes the same material, but narratively, rather than systematically, as above
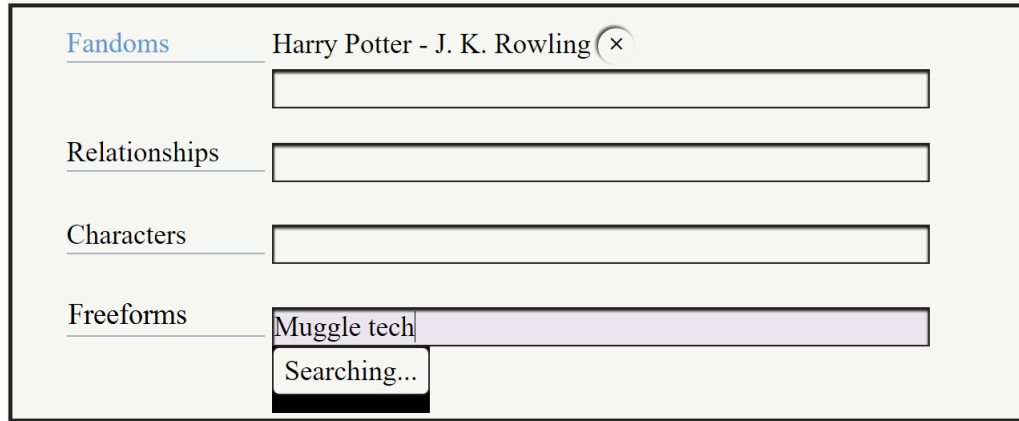
Curated Folksonomy Process at TFR

The following example illustrates the interaction of categories and tag states at TFR, highlighting the wranglers' tasks in creating links between tags.

Posting a New Work and Initiating a New Freeform Tag (User)

- An author posts a fanwork to TFR, entering "Harry Potter – J.K. Rowling" to the fandom tag space and "Muggle tech" in the freeform tag space, as in Figure 3. All of the fields in this entry form and in all others trigger an

autocomplete field as the user types, as shown for the Freeforms field in Figure 3.

Figure 3 Entering tags for a new fanwork

Wrangling the New Freeform Tag into the Specified Fandom (Wrangler)

- The tag appears in the Freeform Unwrangled queue (or "bin") for the wranglers responsible for the "Harry Potter – J.K. Rowling" section of the curated folksonomy, otherwise known as the fandom, as in Figure 4.

Figure 4 Unwrangled bin

- The tag edit page suggests that the Fandom for this tag is "Harry Potter – J.K. Rowling," populated from the user's choice of fandom tag on the associated fanwork.

- A wrangler for this fandom begins writing "Harry Potter" into Fandom field; "Harry Potter – J.K. Rowling" appears in Autocomplete, and they[1] select it as in Figure 5. Once the wrangler saves this change, the tag moves into the Freeform Unfilterable bin for this fandom.
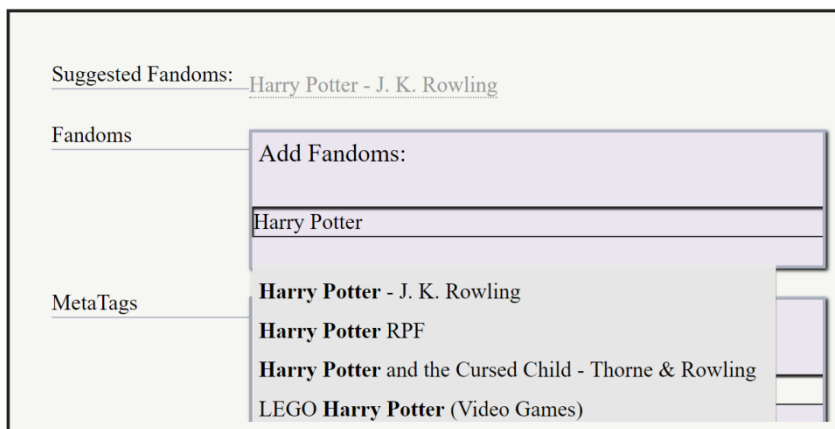
Figure 5 Wrangling to a Fandom

- If the wrangler had determined the tag belonged to a fandom other than "Harry Potter – J.K. Rowling," they could enter that fandom name into the Fandom field instead, moving the tag to the Unfilterable bin for the other fandom. The wrangler would then write a comment on the tag which would send a notification to the wrangler for the other fandom.

Linking the New Freeform Tag to Existing Tags in the Fandom (Wrangler)

- In the "Synonym of" field, a wrangler enters "Muggle Technology," as in Figure 6 This step removes the new tag "muggle tech" from the Unfilterable bin and makes it a Synonym tag.

---

[1] I use the singular "they" to refer to generic wranglers throughout the dissertation. While "he or she" is typically appropriate when referring to individuals of unknown gender, the TFR community explicitly recognizes non-binary genders, necessitating this phrase become "he, she, or they," which I have shortened to "they."
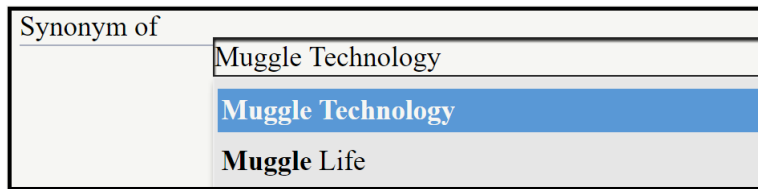
Figure 6 Making an Unfilterable tag a Synonym tag

- Clicking on "Muggle tech" now redirects to a list of works labelled "Muggle Technology," which includes the author's newly posted fanwork.

- In the Metatag field for "Muggle Technology," a wrangler enters the existing Canonical tag, "Muggles," as in Figure 7.
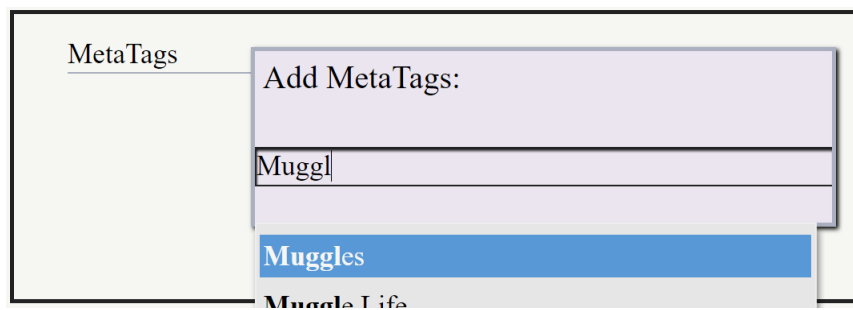
Figure 7 Adding a Metatag to a Canonical tag

- Clicking on "Muggles" takes to user to a page listing works labelled "Muggles," including those tagged with the Canonical "Muggle Technology" and its synonyms, as shown in Figure 8.
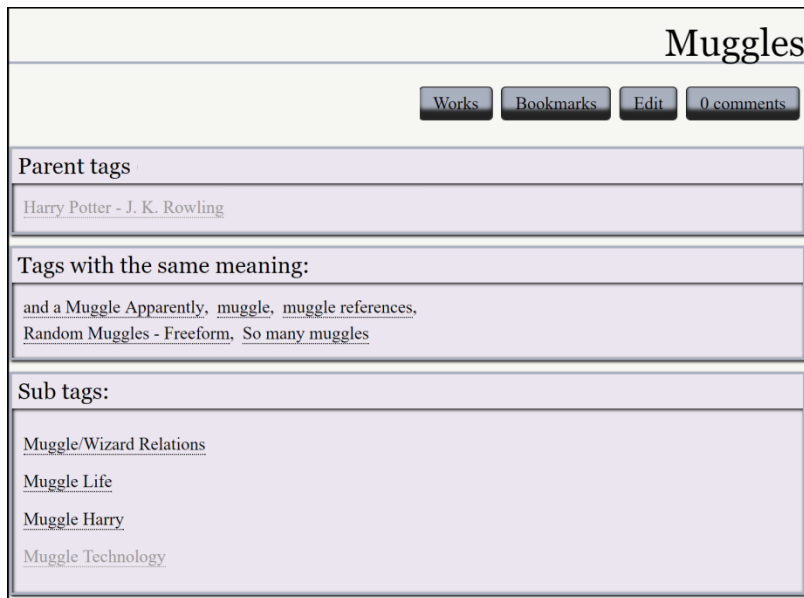
Figure 8 User view of tag page

Adding a Bookmark and Initiating a New Unsorted Tag (Second User)

- A second user bookmarks a different fanwork which helps the user manage a list of fanworks to read or recommend. As part of the action of bookmarking a fanwork, they have the option to add bookmark tags. The user adds the tag "so much muggle tech."

Categorizing an Unsorted Tag (Second Wrangler)

- The bookmark tag appears in the Unsorted tag bin that covers all fandoms, and a wrangler in charge of that bin edits the tag to categorize it as a Freeform tag, as shown in Figure 9.
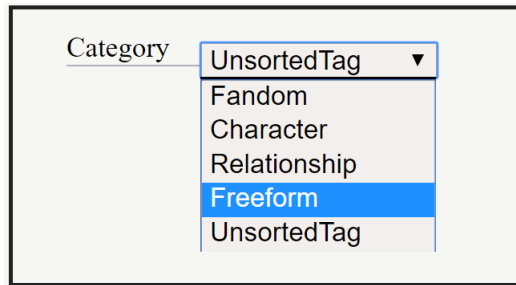
Figure 9 Categorizing an Unsorted tag

- The wrangler then enters "Harry Potter – J.K. Rowling" into the fandom field for the tag, the tag moves to the Unfilterable Freeform bin for the Harry Potter wrangling team, and the process continues as above.

These two accounts—the narrative account and the systematic account—are accurate in terms of the movement of tags throughout the system but simplified in terms of the experience of the wrangler. These accounts do not explain how the wranglers made the decisions that changed the tag: when to address a tag in each queue, and which values to enter into the various fields on the tag's edit page—Fandom, Synonym of, and Metatag. In the following section, Data Collection, I describe the processes by which I learned how the system worked and sought the story beyond these simple mechanics.

DATA COLLECTION

In these subsections, I document how I gained entry to TFR as a participant observer and later engaged in direct data collection. As a participant observer, I gained access to passively collected data sources such as mailing lists and persistent chatrooms. After establishing myself in the community, I used the internal mailing list to recruit wranglers into a diary and interview study, allowing active data collection. These multiple sources of data triangulate my findings (Creswell & Miller, 2000): participants provided me with rationale and histories for phenomena visible to me through my

participation observation, while system data verified the accounts of my participants and made concrete the impacts of their choices. The following subsections are arranged in chronological order reflecting the stages of my engagement with the community and the sequence of my data collection methods. First, I describe how I arranged entry to the community.

Entre & Permissions

I applied to be a TFR wrangler with the understanding that I was joining to perform participant observation research. The chairs of the wrangling committee admitted me to the group in October 2012 with the agreement that, as a participant observer, I would meet the codes of conduct and obligations common to all wranglers. In October, 2014 I re-established this agreement with new committee chairs. The chairs offered their support in contacting and recruiting wranglers in the project for interviews. In addition to my communication with the committee and its chairs, I developed a key informant among the wranglers. This wrangler provided feedback on the first draft of research protocols with regard to the culture and sensitivities of wranglers. This feedback served as a form of member checking, through which I verified my early interpretations with a community insider (Creswell & Miller, 2000).

Additional committees within the organization—those handling the organization's legal and internal communication needs—reviewed my research protocols and offered amendments as needed. The committees requested two changes to meet internal communication policies: to include a statement that this was my own research and not on behalf of the parent organization, and to modify one prompt that asked wranglers about their peers' work. Specifically, to avoid the appearance of conflict with a communication policy not to discuss peers' work in public, the prompt that originally read:

"What would you like to see in tags done by other wranglers?"

became

"If a wrangler asked you to check over their work for them, what would you look for?"

This change put the focus on the wranglers' criteria for good work instead of on prescriptions for others' work, a change that was amenable to the organization and me.

In the following two subsections, I provide detail on how I have used pseudonyms to represent participants and how my participant observer status provided access to and insight into the community.

Pseudonyms

I disclosed to participants that, although I use a pseudonym for the organization, community members are likely to recognize TFR by its description. I took additional measures to ensure fair treatment, trust, and confidentiality in the analysis of participant data. Specifically, in cases where a participant's fandom had typically had a single wrangler, I removed identifying details of that fandom, such as truncating character names, in order to avoid identifying the participant to their peers. Similarly, the pseudonyms I have assigned to participants do not necessarily reflect the participants' ethnicities or countries of origin. Finally, in one instance, I used two different pseudonyms for one participant, since associating two of their fandoms with a single individual would be sufficient to identify them to their peers.

Indicating the gender of my participants was important in terms of accurate representation and in protecting their identities. Only one of my participants indicated he/him pronouns. Most indicated she/her pronouns in their introductions and two used they/them pronouns. Of those I quote and refer to directly, all but one used she/her

pronouns or otherwise presented themselves as female, while one used they/them pronouns. The use of they/them pronouns in this community was rare enough that combining this detail with any other characteristics of the participant (e.g., their tenure with the organization or the type of fandom they were responsible for) could be enough to identify them to their peers. For this reason, I use she/her in references to this participant, with their permission.

Participant Observation Duties

My first duties to the organization included assigning myself fandoms from the list of those currently without wrangler oversight. Over the course of my involvement with the organization, I assigned myself fandoms either as they were created alongside new source material (e.g., *Bioshock Infinite*, a 2013 video game) or as other wranglers unassigned themselves. My largest domain with the most associated fanworks and requiring the most maintenance was the video game series Mass Effect, which I managed first as a solo wrangler and then together with a co-wrangler beginning in June 2013. To extend my range of experience, I joined the *Teen Wolf* wrangling team, one of the largest wrangling teams at 6-8 volunteers, in January 2013.

My three years of participant observation with TFR developed my contextual understanding of the work of these classification designers. I was involved in the community through multiple cycles of recruitment, with new wranglers being inducted, trained, and integrated into existing projects. I was involved in the community under different committee chairs, staff leadership, and their projects to improve the volunteer team's work. Just as deep immersion in the literature prepared me to understand classification theory and design, prolonged engagement in the field yielded a sense of the organization and volunteer team's culture, including indoctrination into norms and

practices, day-to-day experience with peer communication, familiarity with the jargon, and exposure to a variety of events, including yearly cycles of engagement (Creswell & Miller, 2000). Most importantly, my long tenure with the organization has helped me identify the plurality of experiences among wranglers, so that, after confirming and exploring my own experiences with other wranglers, I proceeded with further data collection to hear from wranglers with a different experience than my own.

Joining TFR as a wrangler gave me access to several sources of information not available to the public or to users of TFR: a general volunteer mailing list, a mailing list specifically for the wranglers, a set of persistent chatrooms for all volunteers and for wranglers, hidden system data regarding the tags and wranglers, and the organization's internal wiki. I used chatroom data to familiarize myself with the kinds of difficult questions wranglers pose to each other, and this familiarity figured into the development of my diary and interview protocols. Because wranglers agree that the chatroom is not a completely inclusive communication channel—it excludes wranglers who cannot participate in a synchronous conversation due to scheduling—all decisions made there are also recorded elsewhere, such as in system data or the internal wiki. Therefore, I performed my analysis and took all quotes regarding these decisions from these more formal communication channels. Three subsections below describe the mailing list, the system data, and the internal wiki in detail. These sources figured centrally in analysis and are quoted in the findings chapters.

Wrangler Data Channels

I gathered data through four sources available to me as a wrangler. I saved mailing list messages and chatroom transcripts over the three years I was a wrangler on the system. These messages and transcripts served as records of discussion among

wranglers. Mailing list messages served as a primary data source, and I describe them in more detail below. First, as a participant observer, I took notes on my own process of learning the system, both in familiarizing myself with my first fandoms and in my experience of the daily maintenance of these fandoms. These notes served as a record of how I familiarized myself with the system, its features, and its constraints; they have also aided my interpretation of data from wrangler communication channels. In this manner, the notes provided insights for the construction of my diary and interview protocols. Finally, I subscribed to some optional channels of communication within TFR, such as an email group for discussion of religion and folklore terminology and a second email group regarding concerns specific to music fandoms. These additional channels provided records of more in-depth discussions on these two topics. The following three subsections provide more detail on the wrangler communication channels that figure in my analysis: the wrangling mailing list data, the internal wiki data, and the system data.

Wrangler mailing list data

All wranglers subscribed to a common mailing list and staff encouraged wranglers to use it as a resource for making difficult classification decisions, developing new approaches and guidelines for classification, and coordinating work among the volunteer team. Over my tenure with the team, the mailing list generated 1,624 threads. As a representative example of activity, in July 2013, wranglers generated 44 threads in the list. Most threads consisted of six or fewer messages with three or fewer participants, while a few exceeded 10 messages and one consists of 52 messages.

Of the 44 threads in July, 2013, six reflected coordination of work, primarily wranglers sharing lists of fandoms needing work: new fandoms to the website, fandoms that were previously the responsibility of a retired wrangler, or fandoms that are growing

too quickly for a single wrangler to manage. Eighteen were translation requests in which a wrangler copies tags belonging to their domains but in languages and often alphabets they cannot read. Nine of the threads were staff announcements regarding the technical status of the archive (e.g., that the wrangling interface would be shut down for 12 hours), periodic updates from the parent organization (e.g., reminders of internal wiki policy), and announcements of newly published training documents or guideline change proposals.

The remaining 11 threads were discussions of wrangling itself, and were of primary concern for analysis. These threads include discussions regarding existing or new guidelines for classification work or discussions specific to individual tags. Instances of the latter case included a wrangler's announcement to peers that a tag has been created so that other tags might be tied to it and a request for advice on handling a particularly difficult tag decision. Several of these requests for help were resolved in a single reply, either by identifying an existing solution or referencing a guideline that necessitates a particular outcome. One thread in this set was a discussion of content that was not easily solved but instead produced an extended discussion among 26 wranglers spanning 52 messages.

Beginning in July 2015, the design team brought a systematic review of guidelines to the mailing list. On an average of once a month, staff introduced a new guideline revision topic and moderated discussion in the following order: Questions/Brainstorming, Compromise/Drafting, Conclusion/Implementation. These discussions were popular topics among wranglers and received between 20 and 80 messages each. Before July 2015, mailing list discussions about proposed guideline changes were unstructured conversations.

Internal wiki data

I used three types of documents from the internal wiki in my analysis: tag wrangling meeting notes, fandom-specific wrangling guidelines, and training materials. Wrangling meeting notes were formal documents reflecting meetings the wrangling committee conducted in the chatroom. They provided evidence of the timing of policy and technical changes. Fandom-specific wrangling guidelines and training materials were written collaboratively by wranglers and were subject to regular revision by the group. Wranglers prompted minor changes to the training documents for clarification or correction while the wrangling team developed more substantial changes through open discussion among all wranglers. These documents represented the consensus of the team—all wranglers for the training materials, a fandom cowrangling team for fandom-specific wrangling guidelines—on best practices and what they believed was most important for new wranglers to know before interacting with the live system. In particular, these documents represented the opinions of the experienced, engaged wranglers who contributed to revisions.

System data

My position as a wrangler provided me with access to exclusive system data including comments left on tags by wranglers and liaison volunteers. There were approximately 170,000 threads of tag comments as of December 2016. Most tag comments had no replies, and were wranglers indicating that a tag had been moved to another wranglers' domain. Those of interest for analysis here were discussions among volunteers, such as among wranglers with shared authority over a tag or between a Support liaison and a wrangler regarding a ticket initiated by a user.

Tag comments were not searchable. They occur in my data where their presence was indicated in another source of data, such as a participant mentioning a tag that

generated some discussion. I saved offline copies of tag comments which I identified as relevant. For this reason, and unlike the other types of wrangler communication channels, collecting system data followed rather than preceded participant recruitment. In the following subsection, I describe the process of recruiting participants for diary and interview studies.

Recruitment of Participants

I recruited diary and interview participants in two rounds in 2015. I based my recruitment approach on two related principles: theoretical saturation and coverage. Theoretical saturation is the point at which, in qualitative data analysis, no new concepts or categories emerge from the data (Morse, 2004). The goal of theoretical saturation was particularly important to my aim to thoroughly document the breadth of factors relevant to classification design decisions. Through data analysis—which I describe in more detail at the end of this chapter—I confirmed that my data reached theoretical saturation through the second round of recruitment. A challenge to confirming theoretical saturation was ensuring that my data represented the diversity of perspectives among TFR wranglers. In other words, I used the principle of coverage to confirm that the convergence of concepts in my data was representative of the experience of the entire design team and not a specific domain within it. Initial analysis indicated that the greatest difference in participants' experiences (and therefore the concepts I identified in analysis) was in their fandom coverage, with length of experience in the organization making a secondary difference. For this reason, I did not close recruitment until I had reached individuals representing fandom types participants identified as distinctly complex (for example, sports fandom) as well as participants who had just completed training and founding members of the team.

I first sent out an email recruiting wranglers for participation in a diary and interview study in June 2015. Twenty wranglers responded, of whom 17 completed consent forms. Of the 17, 12 completed the diary and interview study immediately, three delayed their participation until the fall, and three dropped out before completing the diary or interview. One of the delayed participants completed their diary and interview in September 2015. I sent out a second recruitment message in November 2015. Five wranglers responded, of whom four completed consent forms, the diary, and interview. The two participants who had delayed their participation from the summer recruitment phase also completed their diary and interview studies at this time. A total of 19 participants completed the diary and interview studies.

In communication about recruitment I made clear that the study was not a TFR project nor would data be reported back to the organization. By distancing the study from the organization, I hoped to encourage participants to share imperfect and unedited classification practices in their accounts of classification work. This tactic may also have had a biasing effect on recruitment. That is, wranglers who were intrinsically interested in classification procedures may have been more likely to respond while those who volunteer out of a sense of duty to the organization would not have seen participation as an element of that duty. Furthermore, though I explained to wranglers that I sought to capture representative accounts of wrangling work, wranglers with lighter workloads may have felt that it was inappropriate for them to participate as their activity did not match the expectations of the diary study. Participants in the diary study who expressed this concern are included, although others may have had the same concern and declined to contact me. For that reason, I expect that the participants in my diary study best represented the more active and reflective members of the volunteer team.

Diary Studies

Given that my intent was to capture an accurate account of daily classification work, I sought a mode of data collection that would occur as close as possible to the moment of work itself, that would cover a representative period of a wrangler's work, and that would produce concrete examples that might be interrogated in greater detail. Specifically, I sought to minimize retrospective bias that could produce rationalized and sanitized accounts of typical classification work. As a "bottom-up" and temporally close data collection method (Iida, Shrout, Laurenceau, & Bolger, 2012), diary studies are particularly suited to these needs. In the next three subsections, I detail the diary protocol design, the participants' completion process, and the finished diaries.

Design

I designed a two week diary protocol that combines the features of a feedback study and an elicitation study (Carter & Mankoff, 2005). As in a feedback study, participants answered detailed questions about their work as close as possible to the time the work occurred, such as the timing, duration, intent, and notable features of wrangling session. As in an elicitation study, I also instructed participants to record their work in a minimal form, such that there would be enough detail to recall the issue (such as the name of a difficult tag) without requiring them to stop and explain the complexities of the situation at that time. The diary protocol appears in Appendix A.

The diary protocol I distributed to participants is an event-contingent protocol. Event-contingent protocols are ideal for capturing each instance of an event, particularly when that event does not conform to a predictable schedule (Iida et al., 2012; Ohly, Sonnentag, Niessen, & Zapf, 2010). I instructed participants to complete the form for each session of wrangling tasks, with an understanding that participants could make

adjustments for styles of wrangling that were not conducive to this pattern (e.g., several short periods of work throughout the day).

Diary studies carry two considerable drawbacks: difficulty and burden (Carter & Mankoff, 2005; Iida et al., 2012; Ohly et al., 2010). They carry a risk common to any protocol participants complete independently, namely that the participants might misunderstand a difficult protocol and fill it out incorrectly. One way to alleviate potential difficulty is to provide a sample, instructional diary to participants. Doing so risks participants conforming to the example, which is counter to the goal of capturing participants' diverse experiences and understandings of their work. I took an opposite approach in designing a diary protocol with few constraints outside of a few factual questions (e.g., "What time is it?"). The protocol was open-ended and allows for participant interpretation. For example, the instruction to record "an example of a complex wrangling task" did not define "complex," but instead allowed participants to define this term for themselves. Nina provided the following examples of wrangling tasks that serve to show how participants differentiated complex ones from straightforward ones:

> One of these definitely looks like an OC [Original Character], and neither the [fandom] wiki nor the google [search] seem to know who she is. ...okay well apparently what has happened here is that the character named in the story (Joanna Mills) is not the one tagged for (Joanna Wandrey). Wandrey is the last name of one of the actresses of one of the characters mentioned. What even. Unfortunately, "Joanna Mills" is an existing tag in SPN [Supernatural] fandom. Consequently, I'll need to leave this for now and consult Staff/potentially the SPN wranglers later.

> Thankfully, the next character is a straightforward wrangle—Samantha Collins, no existing tag, synned [made a synonym] to new tag Samantha Drew Collins to keep consistency across the fandom. Metatags added, and we move on. (Nina, Diary)

I understood the first example of multiple Joanna's to be the participant's example of complex, while the latter example of Samantha to be straightforward. Other participants, such as Kerri, labelled their examples:

> Easy Task: Went through a half dozen fandom syns [synonyms] listed on Avengers megafandom page (link) to check them for loose tags. Those tags were repatriated to MCU [Marvel Cinematic Universe] or synned [made synonymous].

> Complex Task: Someone wrote a story casting bandom figures as the characters from the video game Corpse Party. This required first figuring out what Corpse Party was, then dividing the tags up by fandom, repatriating to the right places, and leaving appropriate comments to all parties. Exhausting. (Kerri, Diary)

My approach placed the burden on me to interpret diverse participant responses. In addition, I made myself available to participants to clarify the diary instructions. The only questions I received from participants regarded the constraints of the event-contingent design: Although I allowed participants to define "wrangling session" as was appropriate to them, some participants wanted to confirm that the diary protocol would be appropriate when they would not do wrangling work every day, while another wanted to confirm that they could adapt the protocol to suit their schedule of several small wrangling sessions throughout the day.

Completion Process

To avoid unduly interfering with the examined activity and to reduce participant burden and attrition, I encouraged participants to complete their diaries in whatever format was most convenient to them. This decision, as with the flexible design of the questions themselves, shifted some burden from the participants to me as a researcher. For example, the diary instructions included suggestions to complete the diary protocol within the Word document, take notes in another medium, or provide any combination of file types (e.g., notes, images, screenshots) that illustrated their activity. After I received

the diaries, I transcribed audio diaries, typed out handwritten diaries, and combined file types into a single PDF diary for each participant. Participants contacted me to confirm the date they began the diary, and I contacted participants when they were nearing the end of their two-week period to arrange for an interview. A log of participant contact appears in Appendix C.

Finished diaries

Most diaries were text only. However, one participant completed one of her entries as an audio file in a talk-aloud format, and six diaries included screenshots. Additionally, six participants sent additional materials after the interview, including spreadsheet files, screenshots, and photographs of their handwritten record-keeping. Table 2 summarizes the completed diaries including the format (e.g., text only, text with screenshots), any follow-up materials (e.g., screenshots), active days (how many of the 14-15 days included in the diary period had entries), and active minutes (the total number of minutes the participant spent doing classification design during this period). As Table indicates, on average the diaries included sessions from 7 days (minimum 2, maximum 15) or 747 minutes (minimum 60, maximum 2470) of classification activity.

	Diary Format	Follow-up	Active Days	Active Minutes
1	Text only		4	80
2	Text only	Screenshots	4	428
3	Text with screenshots	Photo of paper notes	15	1313
4	Text with screenshots	Text file	5	153
5	Text, screenshots, and audio	Screenshots + photo of desk	6	1173
6	Text only		3	306
7	Text only		15	1062
8	Text only	Screenshots	3	251
9	Text only		3	130
10	Text only		4	324
11	Text only	Spreadsheet	11	2470
12	Handwritten with screenshots		6	1028
13	Text only		6	518
14	Text with screenshots		10	912
15	Text only		10	944
16	Text with screenshots		3	531
17	Text only		5	413
18	Text only		11	2104
19	Text only		2	60
		Average	7	747

Table 2: Summary of completed diaries

Interviews

All participants who completed diaries also completed a follow-up interview conducted as close as possible to the end of the diary period as schedules allowed. I

scheduled interviews at the convenience of participants. Interviews followed a semi-structured interview protocol featuring several general questions about the participants' goals and purposes in regards to the classification system, such as "How important would you say wrangling is to the functioning of [TFR]?" The semi-structured interview protocol allowed me to focus on particular questions while adapting the interview to the responses of the participant (Bernard, 2006, p. 212; Rubin & Rubin, 2005). For example details from the diaries sometimes answered questions listed later in the protocol. Conversely, in interviews with some participants, the questions alone were not sufficient to draw out detail, so that prompts helped stimulate more complete answers. The semi-structured interview protocol appears in Appendix B.

As part of my recruitment process, I advertised to the wranglers an open version of my interview protocol the week of the first recruitment announcement. With help from the wrangling committee chairs with scheduling a time when the chatroom was active but not used for specific purposes, I made myself available for two hours to answer questions about my study and to conduct a sample interview with others in the chatroom. Because chatroom logs are available to wranglers, this transcript was available for potential participants to review.

In the following two subsections, I describe how diaries informed interview protocols and summarize the interviews.

Generating interview protocols

Data from the diaries served two purposes: first, to provide contextual information and examples for use during the interview study; and second, to provide a representation of typical classification work. I read each diary twice, once as it was received and once on the day of the interview. Interviews were scheduled as close as possible to the

completion of the diary so that events were still familiar to the participant. Reading the diary generated questions for the interview and provided an opportunity to review referenced items, such as tag comments in the system and mailing list messages. This synthesis of diary and interview techniques produces more incident-rich accounts than interviews alone (McKenzie, 2003; Zimmerman & Wieder, 1977). As a result of this combination of diary and interview methods, participants answered most of the scripted questions from the interview protocol as they expanded on concrete examples from the diary. For example, this exchange from Silvia's interview begins with an observation from the diary:

> Interview: Just came across something else in the diary I wanted to ask about - it looks like you were going to but then didn't add "Julie [Last Name]" to the Julie metatag. You crossed it out. Why was that?
>
> Silvia: Ok just checked I'm pretty sure at the time it wasn't letting me add it to the metatag. The servers were acting up a bit for me that night but that was the only one it wouldn't let me do no matter how much I reloaded the page. Thanks for reminding me though. I just did it. (Silvia, Interview)

This example gave context to a straightforward wrangling task. Especially in the case of limitations and inconveniences with the technical interface, diaries surfaced details of everyday work that might otherwise be excised from purely recollection-based accounts.

The second purpose of the diary was to provide a representation of typical classification work. Procedural questions from the diary helped to capture the frequency and duration of classification design sessions. My review of participant responses to the open-ended questions also generated lists of straightforward and complex tag management tasks, information resources consulted, and less commonly, lists of record-keeping practices and communication with peer classification wranglers. Given the variety of domains the participants cover, the diaries provide an overview of how wranglers classify their tasks as straightforward and complex.

112

Completed Interviews

A summary of the medium of interviews follows in Table 3.

Interview Medium	# of Participants
VoIP Audio	11
Synchronous chat	6
Telephone	1
In person	1

Table 3: Counts of participants by interview medium

Interviews ran between 90 and 120 minutes. Interviews on synchronous chat were most likely to run the full two hours because of the slower speed of typing and exchange as compared to speech, though participant answers tended to be more complete than in spoken interviews. In the course of every interview, including the in-person interview, participants loaded their classification interface, referred back to relevant system data, and provided links to referenced tags and discussions. At the close of each interview, I invited participants to provide screenshots if they later found an example of a complex design task. Three participants sent screenshots within a month of the interview.

I transcribed the audio interviews. The data for each participant includes their completed diary, interview transcript, and supplemental information including referenced and linked documents as well as follow-up documents. Figure 10 illustrates the timeline of my data collection methods.
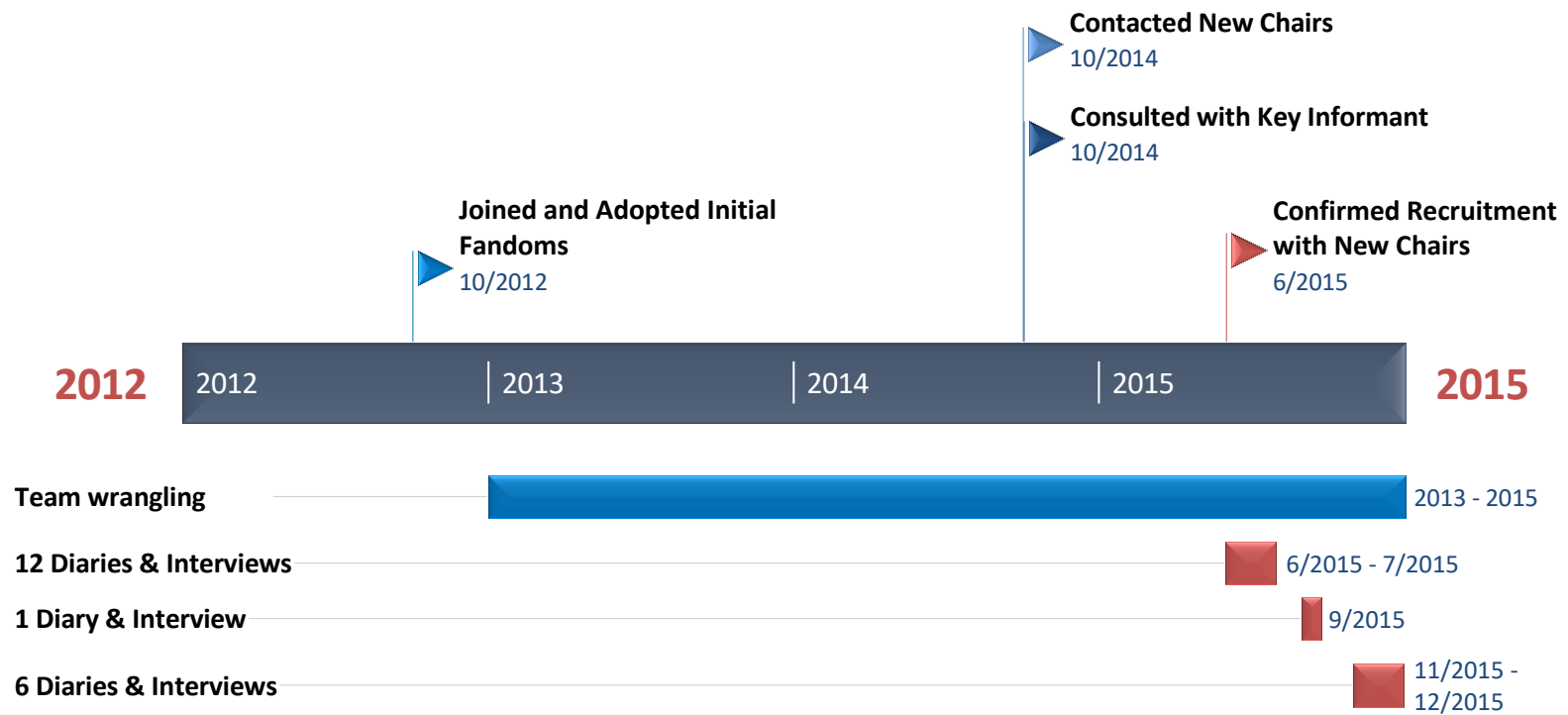
Figure 10 Timeline of participation & data collection

DATA ANALYSIS

I drew on all above sources of data for analysis, with a particular focus on mailing list discussions, diaries, and interviews. I supplemented these sources with relevant system data.

Analysis proceeded according to the tenets of grounded theory (Strauss & Corbin, 1998), though constrained by the aims of this project: to turn greater attention to the process of classification design work. This intention shaped which concepts I found salient at each stage of analysis, though I endeavored to follow the data rather than attempt to impose an external logic, such as the categories and processes established in professional classification literature.

In the first phase—open coding—I reviewed my data to learn how participants describe their own work. At this stage of coding, I used the participants' own language in noting the characteristics of the tags mentioned (e.g., "Relationship tags," "Unwrangled bins," "Metatagging"), participants' interpretation of the tags (e.g., "Misspellings," "Crossover issues"), participants' actions (e.g., "Synning to existing canonical," "Documenting decisions"), and participants' resources or rationale (e.g., "Rule of 3," "Fandom wiki," "Server indexing burden"). Using the constant comparative method (Glaser & Strauss, 1967), I continually returned to earlier instances of codes to determine whether each code was internally consistent or should be divided into new codes. Throughout the analysis process, I wrote memos to document initial observations about emergent patterns and to add contextual information to specific examples.

Through open coding, I determined my unit of analysis to be the classification design decision. In selective coding, I focused on what decisions participants made, how they made them, and how they understood the decisions' relevance to the overall project.

I defined classification design decisions as instances in which participants made a choice affecting the classification system. In my data, these instances included classification decisions as they are typically understood—choices of preferred terms and term relationships, primarily—as well as instances of choices to defer changes to the system and choices of how to make future changes to the system, namely through input to the guidelines of the classification process. I did not include in my definition of decisions those choices regarding the participants' management of their own involvement with the system: the choice of when to begin and end their work sessions. These are interesting questions that I intend to address elsewhere (e.g., Bullard, 2016) and are beyond the scope of this dissertation.

Subsequent passes through my data produced many examples of classification decisions. I further limited my analysis to instances of complex rather than straightforward design decisions. As I noted above, participants defined "complex" through examples in their diaries and interviews; I did not provide a definition nor an example of what constitutes a complex task. By allowing participants to define "complex," I provided space for participants to identify types of complexity other than those I expected and could capture in a definition. In contrast, some "straightforward" decisions may have appeared complex to outsiders (for example, the alphabetization of relationship tags involving transliterated character names) but were second nature to the participants. Analysis revealed that the decisions the participants identified as complex were those that engaged the participants' reasoning and creativity, arose from conflicts within the system, and were not easily resolved by the provided documentation.

Focusing on complex decisions helped me connect the daily work of classification design to broader themes. In infrastructure studies, these instances are breakdowns, through which the typically invisible, taken-for-granted infrastructure becomes visible

(Bowker et al., 2010; Edwards, 2004; Ribes & Lee, 2010) and are part of the broader methodological approach of "infrastructural inversion" (Star & Ruhleder, 1996). In other words, instances in which wranglers experienced challenges helped to illustrate how wranglers understand the system and their rol e in it, especially since the wranglers would then explain to themselves, each other, or to me what the guidelines or processes causing the trouble were meant to achieve.Through axial coding, in which I reviewed the relationships among codes, identified 10 factors that classification designers considered in complex decisions. Furthermore, I discovered that two categories of tags—Character tags and Fandom tags—included instances of all 10 factors that characterized complex decisions. By that I mean, in determining how to handle names (Character tags) and works (Fandom tags), wranglers described factors from seeking multiple sources of information, to diverging from guidelines, to making trade-offs between retrieval functions and authenticity. In Chapters 4 and 5, I present my findings on the factors of decisions through these two foundational slices of classification design: Names (via character tags) and Fandoms (via fandom tags). In the final phase of analysis, I reviewed my data for contradictory examples to my emergent findings. This search for disconfirming evidence (Creswell & Miller, 2000) helped me to further contextualize my findings. In my findings, I also document what I did not find.

Chapter 6 on Guidelines takes a slice in a different direction. In this chapter, I document my findings on how guidelines figured into complex decisions. Guidelines also figure into the other two findings chapters, but here I turn my focus to how guidelines change within TFR. A critical element of infrastructural inversion is a focus on how the infrastructure itself changes, in contrast to studying ways that infrastructure changes the world around it (Star & Ruhleder, 1996). My prolonged engagement in the field has made it possible to surface how the classification design process at TFR changes over time.

Instances of guidelines were identified by the term "guideline," by the close synonyms as "rule" and "principle," by the community-specific terminology "wrangler discretion" and "the FAQ," and by the phrase "supposed to."

Whereas Chapters 4, 5, and 6 illustrate these factors through examples of the complex decisions, Chapter 7 focuses on the factors themselves. Through Chapter 7, "Summary of Findings," I define the factors by synthesizing the previous examples and demonstrate how the factors interacted in designers' decisions. The following quote from a participant diary illustrates my coding for complex decisions:

> Complex Task: A conversation has been going on for months between the various Star Trek wranglers about the freeform tag "Enterprise", which has been used for the tv show Star Trek: Enterprise, the various incarnations of the ship, and, in a handful of cases, to mean a business. I suggested a tag tree to keep any of the previous uses from being permanently misfiltered, but was told that we are no longer to do that in freeforms to reduce server load, so I bowed out of the conversation as I was pretty much out of ideas. (Misaki, Diary)

Using Atlas.TI, I labelled the above quote with the following codes: Complex Wrangling Task, Freeform Tags, Ambiguity, Filtering, Metatagging, and Server Indexing Burden. The first of these codes, "Complex Wrangling Task," signaled that I should consider it for further analysis. The second code, "Freeform Tags," classified it as a different type of tag than those I focused on for my thematic analysis (ie., Character tags and Fandom tags). The final four codes indicate the factors relevant to the wranglers' decision. All four of these factors were found throughout designers' complex decisions and are among the 10 codes I analyze in depth; "Metatagging" was renamed "Hierarchy" to combine it with the closely related codes, "Subtagging" and "Tag Trees."

I analyzed data from all channels of data collection for all three findings chapters. Each chapter is consistent with data found through other sources, though interviews and mailing lists are quoted most frequently. For example, while Chapter 4, "What's in a

Name?" primarily draws on interview data, the themes are consistent with parallel conversations in mailing lists, my participant observation notes, and documentation in the internal wiki.

What's in a Name?

One of the most pervasive challenges across domains in The Fanwork Repository (TFR) was the choice of preferred terms for individuals. The relative difficulty of deciding on terms in proper names is counterintuitive: We might expect that concepts, with their fuzzy boundaries, would be more difficult to manage than the equivalent of authority name records, which in systems such as the Library of Congress, standardize and disambiguate names. For example, the authority name record for the horror author "King, Stephen, 1947-" includes his pen name "Bachman, Richard" and is separate from the authority name record for "King, Stephen, 1818-1852," a marketing scholar. In this example, we might expect to find classification designers struggling to define Stephen King the author's specialty ("Horror"—the proper term and its relationships to other genres) while "King, Stephen, 1947-" would be a straightforward decision. The reason for this expectation is that people, unlike concepts, come with names already attached. In this section, I explore through a few representative examples the challenges in assigning proper names from across TFR domains. These examples show how people, like concepts, are entangled with each other, consist of fuzzy boundaries, and resist static terminology.

The examples in this chapter illustrate how a common classification design decision—deciding the proper name for an individual—required that classification designers practice creativity and create local rather than universal rules. Furthermore, some of the examples in this section highlight the extent to which classification design decisions were not purely conceptual but were embedded within the technical and social infrastructure of the TFR website. In the first section, "Cape Names," I explore the wranglers' difficulties in addressing the multiple and overlapping names in superhero

stories. In the second section, "Schrödinger's Inquisitor," I detail the challenge one team of cowranglers experienced in managing user-determined character names for a video game fandom. In the third and final section, "Changing Names," I explore how wranglers encountered challenges arising from more mundane issues of naming, such as marriage and adoption.

CAPE NAMES

A cluster of proper name challenges occurred around source material dealing with superheroes and secret identities. "Cape names," in fandom terminology, referred to the names characters use for their superhero (or villain) persona, as opposed to their "street names," the names they use in day-to-day life. For example, in the popular DC universe (DCU), Clark Kent's cape name is "Superman." Victoria discussed how issues around cape names played out in TFR, using a related example from the DCU:

> I really wish that they [users who tag] would use their [street] names and not their cape names because it makes it so complicated. […] In DC there's a lot of legacy characters, many characters using the same codename, and characters using really similar codenames like, I mean Superboy is already a problem because there's Superboy Connor Kent who is the one who appears in Young Justice and then also young Clark Kent is known as Superboy, so Superboy is an already ambiguous tag. And then when you make female Superboy, call them Supergirl, then there's already a Supergirl, so there's just no good way to handle that. (Victoria, Interview)

In Victoria's example, several superhero cape names—some from the DCU source material and some the invention of fans—overlapped and created ambiguity. Cape names presented a challenge to the unambiguity of names in that a single person has (at least) two distinct and correct names, such as "Superman" and "Clark Kent." Even if wranglers agreed that these distinct names referred to a single persona, and did not quibble over whether Clark Kent and Superman should be conflated as a single individual in the system, the convoluted history of these comic universes was such that a cape name might

121

have been worn by multiple individuals over time, and one individual may have taken on multiple cape names. Victoria pointed me towards the cape name "Robin," most familiar to casual fans of the DCU as the alternate identity of Dick Grayson in Batman comics, movies, and television shows[2].

> Victoria: I think that in some cases that's been a solution: making a metatag. The way that cape names work with meta and subtagging is very weird and I really should investigate it more because I understood it at one point but it's very confusing. A lot of the time, in character tags at least, the cape name is subtagged to the character name. Which seems like it's backwards and I don't know why that's the system that was settled on. Because there's five Robins and they've each got their own named canonical and then Robin is subtagged to those. That seems wrong.

> Interviewer: That way people that search the canonical "Robin" you only get the ambiguous ones? But if you click Dick Grayson you'll get Dick Grayson and all the ones that could have maybe been Dick Grayson?

> Victoria: Right. So you could get every tag that could be Dick Grayson. Every work that has Dick Grayson in it, presumably, plus some other ones that are not Dick Grayson. (Victoria, Interview)

The structure Victoria described set the street name as a metatag term to the cape name. Figure 11 illustrates one part of the term hierarchy Victoria described. The orange subtags are the multiple cape names Dick Grayson has used in source material works. Cape names such as "Nightwing" and "Red X" were not merely nicknames equivalent to "Robin" but other unique identities of Dick Grayson, complete with their own costumes and storylines. As Victoria confirmed, this hierarchy meant that, in user retrieval, selecting "Dick Grayson" would return all works users had tagged with any of these four character tags.

---

[2] The Robin Problem is a commonly understood dilemma in DC comics, driven by the long history of Batman stories. Robin, Batman's sidekick, is most often the circus performer Dick Grayson, though the cape name has been taken on by many different individuals, both male and female. In addition, Dick Grayson and other Robin street names have also taken on other cape names, such as Nightwing and Red X.
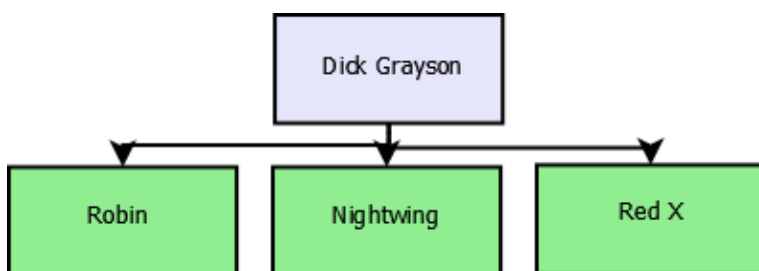
Figure 11 Hierarchy of Dick Grayson character names

Hierarchical relationships such as this one resolved some of the ambiguity of a street name corresponding to multiple cape names. However, due to plot details in the DCU source material, wranglers needed to create a more complex hierarchy. Specifically, Victoria noted that "there's five Robins," meaning that at least one of the cape names in Figure 11 corresponded to multiple street names. That is, just as Dick Grayson has been many superheroes, many men and women have been Robin. Because TFR users tagged for multiple Robin street names from the DCU, wranglers who followed the above hierarchy had to create a polyhierarchy to account for overlapping street name/cape name relationships. In a classic hierarchy, subtags would have only one metatag, though metatags could have many subtags, as in Figure 11. In a polyhierarchy, subtags could have many metatags, as in Figure 12. The example hierarchy in Figure 12 includes more—but not all—of the Robin street names wranglers linked to the cape name. In this more complete hierarchy, "Carrie Kelley" and "Damian Wayne" are also metatags to "Robin," and the additional cape name "Catgirl" is a subtag to "Carrie Kelley." "Robin" is the only cape name shared among the three street names. In Figure 12, the red line points to the subtags linked to the metatag "Carrie Kelley,", the purple line points to the subtag linked to the metatag "Damian Wayne," and the black line points to the subtags linked to the metatag "Dick Grayson."
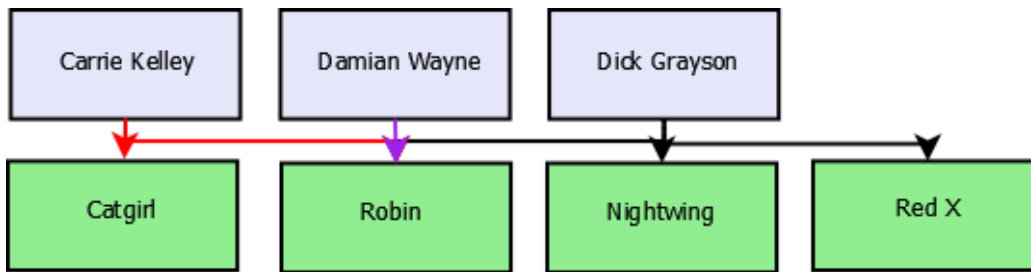
Figure 12 Hierarchy of Robin character names

An even more complete hierarchy would include more street names as metatags to "Robin" and "Catgirl." Each linked street name would add more cape names, and therefore more street names, recursively, until the hierarchy included a substantial fraction of the thousands of characters in the DCU. If wranglers had chosen a reverse relationship—cape names as metatags to street name subtags—the hierarchy would be just as extensive, but would indicate a different relationship between superhero identities and would create different effects for user filtering of fanworks. To understand the factors in the decisions wranglers made in choosing to create—and ultimately dismantle—this hierarchy, I discuss identity and filtering in more detail.

For wranglers, shifting and overlapping identities complicated the typical sense of hierarchies in which classification designers have created relationships between broader- and narrow-terms (general terms for the relationships between meta- and subtags in TFR). For example, in classic hierarchies, such as the Linnaean taxonomy of organisms, terms have an "is a" relationship to terms above them in the hierarchy. "Bats" have a narrower-term relationship to "mammals;" the hierarchy states that "a bat *is a* mammal" but does not state "a mammal *is a* bat." The challenge with regards to cape names in hierarchies was that it made as much sense to say "Dick Grayson *is a* Robin" as it did to say "Robin *is a* Dick Grayson." Throughout TFR, wranglers—and often teams of wranglers, when dealing with very popular source material such as Marvel comics and

DCU—decided among several equally valid term relationship structures for cape names. In the structure Victoria described, above, the wranglers treated Robin as a shared property across multiple individuals, a blurry area in which cape and street identities overlapped. Rather than claim that "Robin" was a class to which all of these individuals belong as a metatag, this classification structure denoted "Robin" as an element of each individual, a subtag. The higher hierarchical position privileged the street identities; it reflected a sense that street identities include superhero identity and that the referent circulating among fans through fanworks was the person who happened to have a hero identity, and not vice versa.

While this structure was consistent with a particular sense of the relationship between superhero and street identities, the specific way that retrieval and metatagging interacted on TFR caused an unforeseen problem. In Figure 12, cape names became the lowest level of subtags, and in the retrieval rules of the TFR system, filtering was inclusive of lower levels of a hierarchy. Therefore, the system always returned the ambiguous cape name whenever users queried any of its specific street name versions, e.g., the system returned all the works containing "Robin" for searches of "Damian Wayne," but searches for "Robin" only returned "Robin" and not its metatags. The rationale for this structure was as follows: if we cannot predict to which street name the cape name "Robin" works refer, all works that refer to "Robin" should be presented to each reader who searches for any of the possible street names for Robin, and the user can then decide—based on other information such as the rest of the tags and the story summary—whether the story is truly relevant to their search. It might even be the case that authors wrote some stories about Robin in such a way that the character could have been any incarnation of the superhero. That is, a user could interpret some stories tagged "Robin" as referring to Dick Grayson, Damian Wayne, or Carrie Kelley. However, the

relationships between street names and cape names were heterogeneous, because individuals such as Dick Grayson had multiple superhero personas. Moreover, variance in fanwork volume across TFR added to the problem: For example, in 2016, there were more than 10,000 works on the archive for "Dick Grayson" and only 4,000 for "Damian Wayne." Having both tags return the 900 works for the ambiguous "Robin" had a much greater relative effect on users' filtering search results for "Damian Wayne" than it did for users searching for "Dick Grayson." That is, nearly 25% of the "Damian Wayne" results were likely to be irrelevant to the search, while only 9% of the "Dick Grayson" results were likely to be irrelevant to that search. In the terminology of information studies, we would describe this situation as a decrease in the precision (Cleverdon & Keen, 1966; van Rijsbergen, 1979) of the "Damian Wayne" search, whereby search results for "Damian Wayne" would contain more irrelevant works as compared to search results for "Dick Grayson." This decrease in the precision of filtering was the substance of a Support ticket, in which a user complained about the number of irrelevant "Robin" works appearing in search for "Damian Wayne." This complaint spurred a change in structure, from subtagging "Robin" to each street name to separating the names entirely. Prompted by a user complaint and in light of the effect on filtering, wranglers chose to undo the hierarchy between Robin's cape and street names.

Based on the concerns about search results and the complexity of identities in the DCU, wranglers agreed that there were competing, valid hierarchy solutions for cape names. As Victoria stated, the hierarchies were confusing even to those within the wrangling team. Though the decision on a given hierarchy was contentious rather than definitive, once made, the hierarchy played an influential role in design decisions for wranglers outside the DCU wrangling team. Decisions made for well-known properties such as Marvel and DCU rippled across TFR as wranglers consulted those fandoms'

canonical tags and hierarchies as authoritative examples. Matilde, a wrangler tackling a new animated series searched through Marvel and DCU canonical tags as part of her decision-making process for characters in her fandom:

> Matilde: I wasn't sure which canonicals to pick. So I went to another superhero universe that I knew and which I would expect people to use different types of tags for characters and just had a look at how that fandom was wrangled to make a better decision for my own fandom. Or at least to help me decide what to do with my own fandom.
>
> Interviewer: So what did you end up going with for [your fandom]?
>
> Matilde: I ended up going with just the normal people names [street names]. […] I ended up deciding to go with just the normal person name for two main reasons. One is because the superhero name is actually used not just for them but also for other characters. Even though in the show they are not major or [do] not just get mentioned. Kind of like in the case of Batman if you know that. The Batman character is not just one person, it's been many different people, so I felt that it would have been problematic to include the hero name if someone would eventually use it for another version of that superhero. Mostly because of that. And also because their identities are known. They don't—in the show, the characters don't know their identity themselves but the viewer immediately knows the secret identity of the character, so it's not spoilerish or anything. (Matilde, Interview)

Matilde was dealing with a relatively new but quickly growing fandom and, based on her knowledge of the show and with reference to established superhero characters such as Batman, she anticipated character tag problems would ensue from overlapping cape names. She went on to confirm that fanworks about the other identities had not yet been added to TFR, so that she was able to make all uses of the cape name synonyms to the street name, but planned to continue to monitor usage in the fandom over time in case a more complicated structure, as above, became necessary. Matilde expected that making the street name the canonical character tag would influence future users to use that name, rather than the cape name, so that fewer ambiguous uses would accrue before a split became necessary. Specifically, this influence would occur through autocomplete, where

canonical character tags appeared in a dropdown list of suggestions as users typed in the character tag field. Shortly after our interview, Matilde created a subtag structure in which the street name was a metatag for the cape name, following inevitable uses of the cape name to mean other instantiations of the hero, similar to tag uses for Batman or Robin.

Matilde provided as her second reason for following the precedent set for canonizing street names that the names were not "spoilerish." By "spoilerish" she meant that revealing through tag relationships that Dick Grayson and Robin are the same person, for example, would not reveal new information to the user. Redirecting a user who searched for "Robin" to "Dick Grayson" would not spoil a revelation later in the series. She did not, however, say whether or not a spoilerish cape name/street name relationship would be sufficient for her to choose a different tag or tag relationships, only that the lack of a spoilerish relationship made her decision easier.

The issues that Victoria and Matilde encountered also occurred in other superhero fandoms, so much so that wranglers included the cape name phenomenon in training sessions for new wranglers. An experienced wrangler summarized the issue with cape names during a character tag training session with new wranglers. In instances where characters have more than one name in the source material, she advised:

Option 1: Pick the most common/clear name and syn the other version to it.

This is the easiest option. If a character mostly goes by a stage name or pen name in all public contexts, pick that if it's unique. Your goal is to choose an unambiguous name that users will easily find in the auto-complete. (Training session notes, Internal wiki)

Though she qualified this rule with a reference to superheroes and cape names:

> Comics: Because superhero names are shared, all superheroes should be canonized by their street names, if they have one (that is, "Kate Bishop," not "Hawkeye"). (Training session notes, Internal wiki)

In this training session, the experienced wrangler separated out superhero names as resisting the "easiest option" because many characters shared these names. Though wranglers developed the policy, as shown in the training excerpt above, to make street names the canonical character tags, the more complex decisions of hierarchy continued to be left up to the fandom wranglers themselves.

Cape names resisted easy classification because they contradicted two assumptions about proper names. First, cape names complicated the common sense assumption that a person's name is unique to them now and will always be unique to them. We are familiar with the chance that, coincidentally, two individuals will have the same name; in Library of Congress authority name records, for example, designers have added years of birth and death to individuals' names to protect against the probability that two individuals will have the same first and last name. In the case of cape names, the additional complication was that the many versions of Robin shared an identity; the convergence on a single name was not a coincidence but a plot device. That is, when a new street identity took on the cape name "Robin," the character did so as an intentional reference to the prior incarnation of the superhero. The referent, and not just the term, was ambiguous and multiple. Second, cape names challenged the assumption that identity is singular and fixed, as they presented an extreme and concrete example of a postmodern approach to personhood, in which identity is multiple and in process (Gergen, 1992). For example, a common trope in superhero stories has been for characters to wrestle with notions of self, including whether any of their identities is authentic and whether they have changed themselves by taking on new names (Coogan, 2009; Smith, 2009). For the

sake of the curated folksonomy, wranglers had to decide which name—and therefore, which partial identity—was true, primary, or inclusive of the others.

In making decisions on canonical character tags for cape name characters, wranglers had to define clear identities from source material works which were inherently about a lack of a clear identity. Since cape name source material defied these two assumptions about proper names—that names are unique and that identities are singular and fixed—wranglers such as Victoria and Matilde could not take the "easiest option" but reflected on multiple factors, including what was authentic, how filtering would be affected, and how their decisions might spoil source material plot for TFR users. That is, they considered multiple purposes of a classification system: those aligning with perspectives such as consensus warrant, user warrant, and ethical warrant. In the next example, wranglers also experienced challenges inherent to the source material though user choices contributed substantially more ambiguity than in the case of cape names.

SCHRÖDINGER'S INQUISITOR

The second example of the challenges in assigning proper names to characters came from scenarios in which readers of a text—in this case, players of a video game—customized character names. Video games with customizable main characters presented a particularly pernicious example of the distance between names and identity. In story-driven games, such as those from the game developer Bioware, the main character's gender, appearance, profession, and first name have default values (e.g., "Jane Shepard" and "John Shepard") but are customizable at game start. In Brigid's diary, she introduced a recurrent challenge to character tag decisions as "Playing what I call the 'Schrödinger's[3]

---

[3] Brigid was referring to a thought experiment devised by Erwin Schrödinger meant to illustrate how some entities in quantum mechanics can simultaneously exist in multiple, contradictory states until observed (Trimmer, 1980).

Inquisitor game.'" In the video game, *Dragon Age*, the main character—"the Inquisitor"—had variant names depending on story choices and user customizations. Since many users indicated the main character by using one of four possible last names, Brigid and her cowranglers encountered ambiguity regarding other uses of those last names:

> Users who pick a different name for their character will often create a character tag for that name, e.g., in today's case Ariana Trevelyan. However, sometimes users will also create Original Characters sharing a last name with the Inquisitor. So Ariana Trevelyan could be a genuine Inquisitor, or could be the Inquisitor's mother, sister, wife, daughter, second cousin once removed, etc. So for something that looks like an Inquisitor and quacks like an Inquisitor, the challenge is to work out whether it really is an Inquisitor, or an Inquisitor-related OC [Original Character][4]. (Brigid, Diary)

The difficulty in determining whether or not a character tag referred to the game's main character or a character that was the invention of the fanwork creator sprung from user tags including last names from the source material.

When Brigid discussed the matter of "Schrödinger's Inquisitor," she was chiefly referring to the question: Was this unambiguously an Original Character not representing the main character from Dragon Age? Tags sometimes included unambiguous information, for which Brigid was very grateful:

> Everything that is very obviously not an Inquisitor—sometimes people will flag, will actually include in the tag, OC [Original Character], and great, I love them when they do that. Or non-Inquisitor, great, love you. [She laughs]. And we'll syn those to Non-Inquisitor Lavellan or Trevelyan or whatever it is. (Brigid, Interview)

Without this unambiguous information in the tag, Brigid and her cowranglers had to determine the referent of the tag from context, such as from other tags for the fanwork

---

[4] Original Characters, in fanworks, were characters invented by the fanwork creator and not found in the original source material. In the case of Bioware games, though the player made many decisions about the main character, the main character was not an Original Character, while an invented sibling would be.

and the content of the fanwork itself. If the wrangler's investigation did produce a clear answer that the character was an original invention of the fanwork creator, such as a possible relative of the game's main character, then the wrangler faced a second question: Should this tag be made synonymous to a "Original Non-Inquisitor" character tag? Overwhelmingly, Brigid and her cowranglers chose not to make synonymous to "Original Non-Inquisitor" character tags those tags that did not include cues such as "(OC)" or "Non-Inquisitor." Of the 20 tags synonymous to the canonical character tag "Original Non-Inquisitor Lavellan"—the largest tag of this type measuring by associated works—only one did not signal in the tag itself that it referred to an Original Character. Two of the other three "Non-Inquisitor" canonical character tags had no ambiguous tags in their synonyms, and for the second-largest tag of this type, Trevelyan, two of 17 were ambiguous. Brigid stated that making these few ambiguous tags synonyms had not produced any problems—namely, complaints or queries from users passed on from Support—but the decision to make them synonymous had a flaw, since another user might coincidentally use one of these same names to refer to the game's main character. For example, though a user first tagged a work with "Llana Lavellan" to indicate an Original Character, a future user might use "Llana Lavellan" to indicate the game's main character. While Brigid had become vigilant not to make tags without details such as "(OC)" synonymous to the "Original Non-Inquisitor" tags, others among the fandom's cowranglers sometimes chose to use all of the information available and made ambiguous tags synonymous—risking the small chance a future user might use them for a different referent.

It was not immediately clear to Brigid's cowrangling team that this ambiguity between main character and original characters existed in the fandom's character tags. The wrangling team first considered the "Original Non-Inquisitor" dilemma through

discussion in the comments for one of the "Original Non-Inquisitor Lavellan" synonymous tags. One wrangler prompted the discussion by stating, "I haven't been synning them [making them synonymous] to the Inquisitor tree. In my head people looking for Male/Female Lavellan are looking for Inquisitor Lavellan" (Internal system data). The impetus for dividing these Original Characters from the hierarchy of versions of the game's main character was to remove these characters from search results for the top-level ambiguous tags such as "Female Lavellan" or "Trevelyan." Previous to their team-wide discussion, the cowranglers had linked character tags to the Inquisitor hierarchy even when the tags did not contain the term "Inquisitor," and the cowrangling team assumed that these tags unambiguously referred to the game's main character. It was not until more than six months after the game's release and more than 5,000 uses of tags in the Inquisitor tree that the wranglers encountered enough tags referring to Original Characters sharing the main character's last name—around 25 at that time—to deem it necessary to distinguish these tags from the Inquisitor tags. Challenges such as Schrödinger's Inquisitor arose over time, as a result of user tagging choices not initially perceived by the wranglers, and became a regular task for wranglers. Even as users themselves, wranglers in this fandom had not predicted the extent to which Original Characters would become a problematic type of character tag. A related challenge arose when the wranglers determined that a character tag referred to the main character, leading to a slightly more complex practice for resolving ambiguity.

In the comment on the "Original Non-Inquisitor Lavellan" tag, the wrangler I quoted above mentioned "the Inquisitor tree," referring to a large hierarchy of canonical character tags linking variations of the main character. Brigid stated there were two main tasks when receiving a new character tag that resembled an Inquisitor name: deciding whether the character tag referred to the main character or an Original Character, and

determining to which canonical character tag it was appropriate to make the tag a synonym. As detailed above, if the outcome of the first task was the determination that the tag referred to an Original Character, the wranglers had to decide whether to make the tag synonymous to an "Original Non-Inquisitor" character tag. If the outcome of the first task was the determination that the tag referred to the main character, the wranglers then had to decide to which canonical tag for a specific version of the main character to make the tag synonymous. Brigid listed three complicating factors to this choice of specificity, each reflecting one of the three story-centric decisions players made about their main character: species (which determined the surname, e.g., Lavellan, Trevelyan), gender (male or female), and profession (e.g., mage, warrior). In the interview, Brigid explained how she and her cowranglers came to create an exhaustive hierarchy to cover each variation users created tags for Inquisitor characters in TFR:

> So we have the Inquisitor tree, and there are four different last names for possible Inquisitors, take Lavellan. We have Male Lavellan, we have Female Lavellan, and then we've got even more specific branches like Female Mage Lavellan and Female Warrior Lavellan et cetera. So what we want to make sure is a) that we distinguish between characters that are intended to be the Inquisitor and characters that are not. But also [b)] that we syn to the right level of the Inquisitor tree. (Brigid, Interview)

The tag hierarchy, or "tree," Brigid described was one of the most complex in the TFR curated folksonomy. Figure 13 illustrates the placement of one of the tags she mentioned, "Female Mage Lavellan." This character tag was at the bottom level of the hierarchy as it was a fully specified variation of the character. Wranglers subtagged it to its two less specific variations, "Mage Lavellan" and "Female Lavellan," which the wranglers in turn subtagged to even less specific variations. In Figure 13, each line represents a metatag-subtag relationship.
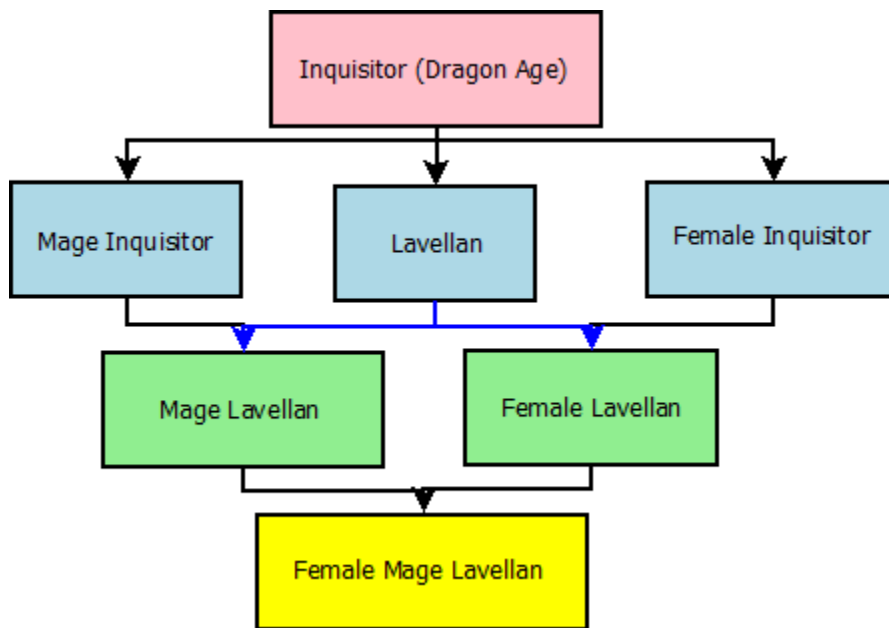
Figure 13 Hierarchical structure of Inquisitor character tags

Because the polyhierarchical structure of the curated folksonomy allowed for subtags to belong to multiple metatags (and vice versa), each branch of the hierarchy in Figure 13 led back to the same "Inquisitor (Dragon Age)" tag. The wranglers built this hierarchy to reflect the variations of the main character possible in the source material and that users had tagged for in TFR. In deciding to make a new tag synonymous to one of these canonical tags, wranglers considered how much specificity was indicated in the tag itself.

Similar to the two-step process for resolving Original Character tags, when the wrangling team encountered tags that did have as their referent the game's main character, they had to make a second decision. Brigid stated that the team had to make sure "that we syn [make synonymous] to the right level of the Inquisitor tree." She went on to explain:

> One of the problems we had when I started wrangling was I assumed that if I knew that this character was a female mage Lavellan that that was the right place

to syn it to because, it was the most specific possible and if people filtered for Lavellan, because it's subtagged, they would get it, but if people just wanted to read Female Mage Lavellan they would get that. The problem is that some of the names that people use as first names are quite common. Obviously, the ones that are preset names in the game. In the case of Lavellan, the male is Mahallen. So lots of people use that. And there are lots of common first names. So if two users use Mahallen Lavellan as once a mage and once a warrior you obviously can't syn it to both.

We decided actually it's best to syn it to the gendered version if you knew what the gender of character was from that first name - if it was very, very obvious or just to the surname if it wasn't blatantly obvious what the gender was. And that's a really tricky one sometimes because if you think about how names get read in different cultures, you can't always assume that just because it's a male name in English it'll be a male name in other languages. Andrea is a classic example. It's a female name in English but it's a male name in Italian. So you've got to be really careful and it's kind of gut feel and can this possibly be read in any other gendered way in which case it goes in the non-gendered. (Brigid, Interview)

Of the three story-relevant choices a player made regarding the game's main character, the game indicated only species (e.g., human or elf) within the character's name. In this excerpt, Brigid noted the temptation to use contextual information to determine the other two major choices—gender and profession—and make tags synonymous appropriately. The risk, as she described, was that attributes knowable through the first use of the tag were not necessarily true for all uses of the tag, especially when players reused first names. Her approach—and the one advocated among her cowranglers—was to err towards the general, rather than the specific. Of the character name tags for Inquisitors, wranglers only made those that included the word "mage" or "rogue" synonymous to tags specifying class. Likewise, they only made those names that were "fairly obviously" gendered male or female synonymous to tags specifying gender. In these decisions, Brigid's team was more consistently conservative in judging ambiguity than in the decisions to make tags synonyms to "Original Non-Inquisitor" tags. In part, this more conservative judgment was due to the wealth of options available to the wranglers; if the

information was not available in the tag itself, it could be made synonymous to a more general level of the tree.

Brigid and her cowranglers balanced the outcome of retrieval deep in the Inquisitor hierarchy against the possibilities inherent in the use of ambiguously named character tags over time. Placing tags lower in the tree using a strong test for specificity prevented future problems arising from users' coincidental reuse of the same name for a different variation of the inquisitor. Because the system only prompted wranglers to review a tag when a user first created the tag, a second use of "Andrea Lavellan" would not appear to wranglers for review, but would inherit the same tag state and relationship as previous uses. If Brigid had continued making tag decisions based on contextual information for each fanwork, the system would have invisibly mis-indexed future works until or unless a user noticed the error and submitted a Support ticket. For Brigid's wrangling team, this danger was greater than the drawback to retrieval in placing tags higher in the Inquisitor tree. Because filtering was only downwards-inclusive, if users selected "Female Lavellan," they would not retrieve works tagged with only "Andrea Lavellan." In part, Brigid and her cowranglers were assuming that users understood the tag system well enough to filter by tags at the appropriate—in this case, the highest— level of the hierarchy. Similarly, their choice rewarded users who created very specific tags, such as "Andrea Lavellan (Female Mage Inquisitor)," as the wranglers would make these tags synonymous to the lowest level of the hierarchy and therefore include those tagged works in more filtered results.

As with cape names, the core challenge of Schrodinger's Inquisitor was the poor fit between a referent (an identity) and a tag (a proper name). In case of inquisitor character tags, the multiplicity of identities in a single name was present (e.g., the ambiguity of "Andrea Lavellan," man/woman, mage/warrior), but the multiplicity of

names for a single identity was astronomically worse than in the case of cape names. Wranglers in this fandom were not struggling to resolve the messy properties of multiple reboots over 80 years of comics; they were struggling to resolve the possibilities of completely user-defined first names across 13,000 works starring the inquisitor character. Dragon Age players—and TFR users—shared a common referent but with few technical or conventional restrictions on its representation in a name. Brigid and her wrangling team made choices that recognized the shared referent, and that the shared referent had meaningful variations according to gender, species, and profession, while cautiously accounting for user tag choices that defied over-defining a single instance of a tag.

In the two examples of cape names and Schrodinger's Inquisitor, challenges arose for extraordinary reasons—the source material works created a phenomenon in which a single identity had many names and vice versa. Outside of comic books and royal dynasties, it is less common for individuals to take on new names and share their names and identities with others. Beyond player-centric video games, it is not common for the reader of a text to decide the name for the main character. However, the extraordinariness of these cases brought into sharp relief naming problems that wranglers encountered across all fandoms. Indeed, any one aspect of the problems wranglers described above— an individual having multiple names, a name referring to multiple individuals, and user choices exacerbating these challenges—would be encountered in any classification system that addresses names and identities. The following subsection details a more familiar, real-life phenomenon complicating wranglers' decisions: that an individual's name changes throughout their lifetime.

CHANGING NAMES

In fandoms featuring either fictional or historical figures, wranglers struggled with how to name characters who underwent name changes. Examples of common name changes included married names in soap operas and in dynasty sagas such as *Game of Thrones*. Examples from across the fandoms that my participants described proved that name changes were not only for superheroes. The guidelines addressed the ubiquity of name changes among characters by pointing to four considerations relevant to wranglers deciding on which name to make the canonical character tag:

> For a character who changes their name over the course of canon (due to adoption, divorce, gender presentation change, marriage, or other plot reason), choose the canonical based on:
>
> - The character's in-canon stated preference;
>
> - Which name they went by at the start of canon;
>
> - How long they went by a name (i.e., were they known as one for longer than the other in the canon);
>
> - Which name the fans most often use/is most common on the Archive.
>
> If multiple names are commonly used in the fandom, the canonical can incorporate them both; e.g., "Lily Evans Potter." If the names are sufficiently distinct, canonicals can be constructed with a "|", similar to translated names (Internal guidelines)

This guideline extended the simple "pick the most common/clear name" instruction I quoted from the training session by pointing to how wranglers should judge what is "common." The four considerations in the guideline included two different sources of evidence—source material (canon) and user tagging choices—for wranglers to use in making decisions reflecting changing names over the course of a character's life. Unlike the issues in the prior two sections, the core challenge for common name changes was not whether character tags share the same referent. Instead, the challenge was to pick which

tag was the best representation of that referent, based on these four—possibly conflicting—considerations.

Two of the considerations wranglers included in the above guideline referred to temporality: the name used first and the name used over the longest span of source material. As Amita explained with an example from her fandom, the timing of source material and fanworks was sometimes more complicated than this guideline suggested, as wranglers made decisions while the source material continued to unfold:

> I picked it up when only the first book was out. And at that point he was known as Four but by the end of the books in the series he's pretty much only ever called Tobias. Someone actually wrote in and commented and was like, "Why is his tag Four? Shouldn't it be Tobias?" And they [staff] were like, "Well, you can do a piped, you have an option here, you can do a piped tag, you can do a Tobias "Four" Eaton tag," and I was like, "Whaaa." Again, by that time usage on [TFR] had definitely tilted towards Four but it bothered me because I knew it was incorrect, technically. Anyway, I just realized, I should just go ahead and bite the bullet and change it to "Tobias | Four" but I've been reluctant because not everyone's been using this other tag. (Amita, Interview)

In this case, fanworks and tag usage on TFR developed contemporaneously with the ongoing release of the source material, a young adult book series. Whereas the guidelines quoted above referred to the source material in retrospect, wranglers initially made correct decisions in this fandom that were later thrown into question by developments in the source material and in new fanworks. Amita went on to describe how diverse experiences of the source material among TFR users made her hesitant to make the suggested change:

> But I don't want to spoil people. This is kind of the same feeling I had honestly when Four's name became Tobias because it is a plot point in the books, you don't know what his name is until later on. I was like, it defeats the whole purpose to have his name be right there. To have his full name spelled out for you if you haven't finished the series. (Amita, Interview)

140

Now that the book series was complete, it was reasonable, according to the guidelines, for Amita use the name "Tobias Eaton" or a variation such as "Tobias 'Four' Eaton," but Amita was sensitive to the inevitability that not all TFR users interested in the fandom had finished the series, especially since a new surge of fandom users were following the ongoing release of the book series' movie adaptations. It was possible for users to avoid stories based on materials they had not read, such as by filtering by date published or by avoiding fanworks authors tagged with tags indicating specific books (e.g., "Allegiant Spoilers"). Amita's choices would have impacted these users not by directing them to works which contained spoilers but by making a spoiler prominently visible in the metadata. Making "Tobias 'Four' Eaton" the canonical character tag would make that complete name appear for users who typed in only "Four (Divergent)"—in autocomplete and as the title of search result pages. In this case, Amita judged that the three considerations referring to source material were inappropriate given that not all TFR users were yet familiar with the complete source material. Instead, Amita made a choice that would make the system more inclusive for users who were following along with the new movie series.

The temporality of source material, fanwork, tag usage, and wrangling presented another complicating factor to the four-point guideline. In the first of the above quotes from Amita, she referenced the proportion of user tags, noting that it had "definitely tilted toward Four." The usage to which Amita referred was the proportion of users who used the canonical tag itself and not one of the variant synonyms. At the time of our interview, the proportion of usage between the canonical tag—"Four (Divergent)"—and its 12 synonyms—all of which contained an element of the revealed name, "Tobias Eaton," was 400:20. The problem with using this proportion of user tagging choices as evidence was that the proportion may have been result of the earlier decision to make "Four

(Divergent)" the canonical tag. In a conversation with Amita, a staff member noted, "Fandom usage – on [TFR] and off is a good guideline, but [TFR] format does influence" (Internal system data). User tagging choices and proportions were valid evidence for choosing among competing names for a canonical character tag but could also be an outcome of wranglers having made a canonical character tag. Influence, in this case, was circular, between user tagging choices and wrangler design choices, as design choices became embedded in the system. Amita, making this judgment years after she made the canonical character tag "Four (Divergent)," was unable to tease apart how much of that proportion was the result of user tagging choices and how much was from her own influence as a wrangler.

A prominent example among Marvel characters further highlights the reciprocal relationship between canonical tags, source material changes, and usage. Fitting with typical comic book and superhero logic, when Skye, a character in the popular television series *Marvel's Agents of S.H.I.E.L.D.*, gained a new name, that new name—in this case, a revealed birth name for the adopted character—happened to be the same name as an existing character in the Marvel comic series, Daisy Johnson. As TFR users began tagging for the television character using "Daisy Johnson," they conflated works about the comic book character and the television character. Wranglers considered one proposal to change "Skye (Agents of S.H.I.E.L.D)" to "Skye | Daisy Johnson" and to make this new tag a subtag to the existing and now ambiguous "Daisy Johnson." This change should have made it possible for television fans to choose the more specific tag as it would now appear in autocomplete when typing "Daisy Johnson."

Users objected to wranglers adding "Daisy Johnson" to "Skye," referring to the guideline regarding in-canon stated preference. At the time when this ambiguity first began to cause a problem on TFR, the character had chosen not to use this birth name.

Users and wranglers commonly invoked this consideration in the guideline to overturn other factors (such as time spent with each name) in the case of name changes around gender identity and transgender characters. Wrangled viewed this consideration as making the system more inclusive by reflecting progressive perspectives on gender. Following this consideration, it would have been appropriate for the wranglers to leave "Daisy Johnson" out of the canonical, as the character had rejected it.

In the case of "Skye" and "Daisy Johnson," the need to disambiguate the two identities in such a way that the site would guide television fans away from the ambiguous term was too urgent for the wranglers to delay adding "Daisy Johnson" to the tag until the character's in-canon stated preference changed. There were two reasons for this urgency. The first reason was the speed at which users were posting new, ambiguously tagged works. Wranglers were motivated to resolve the ambiguity quickly so that users would choose an unambiguous tag for the hundreds of works they were likely to post over the coming weeks. The second, related reason was that the number of "Skye (Agents of S.H.I.E.L.D)" tags was reaching the threshold at which changes would substantially increase the server indexing burden. Specifically, wranglers had previously agreed to schedule changes to tags linked to over 5000 works to TFR's low-traffic times to avoid slowing the server's response to user requests.

Wranglers opted to use "Skye | Daisy Johnson" as the canonical character tag and, as the source material continued, the tag eventually became consistent with in-canon usage, as the character accepted both her adopted name and birth name. Wranglers avoided a hierarchical relationship between "Skye | Daisy Johnson" and "Daisy Johnson," however, for the same reason as in "Robin," above: the relative popularity of the two characters on TFR was such that the results would overwhelmingly return works from the subtag "Skye | Daisy Johnson," making the ambiguous term almost useless for

retrieving comic-Daisy fanworks. The hierarchy, in that case, would disproportionately and negatively affect the smaller group of users searching for comic-Daisy fanworks. In both of these choices—the content of the character tag and its relationship to other tags in the system—wranglers diverged from guidelines to improve filtering: directly, by avoiding a television-dominated hierarchy, and indirectly, by leveraging autocomplete to guide users to a disambiguated version.

Ultimately, the decision of how to handle conflicts between variations of a character name was up to the wranglers responsible for a given fandom. Commonly, wranglers developed guidelines for decisions within a single fandom that extended or diverged from the general guidelines. For example, it would have been within Amita's authority to make a guideline within her fandom that canonical tags should not contain spoilers. However, even within a single fandom, character tags resisted standardization. As Justine, a wrangler for the *A Song of Ice & Fire* series and its television adaptation *Game of Thrones,* noted, differences in the storylines and characterization for married women in her fandom were such that one character was commonly and intuitively known by her maiden and married name ("Catelyn Tully Stark") while another was only ever referred to by her maiden name, even while she was married ("Cersei Lannister"). The complexities of the source material and users tagging according to source material characterizations drove inconsistencies in how wranglers formatted character names.

Justine explained that while internal guidelines were important for formatting character tags correctly, the greatest impact of character name changes was on relationship tags that contained those characters:

> But the really issue isn't the character tags themselves, rather what do we do with the 20 billion relationship tags. Because whenever you've got a character--it's not exponents but factorial I guess. Because with so many characters, you can ship [relationship] them with everybody else! And people do. (Justine, Interview)

Though Justine wanted to push for greater standardization of character names in her fandoms, particularly with regard to how nicknames and married names were handled, she reflected on how this standardization would have to be reflected in the canonical relationship tags containing those characters. In the case of Catelyn Tully Stark, Justine's example, the single character tag was linked to 48 canonical relationship tags, 37 of which wranglers would have had to change to be consistent with new in-fandom name guidelines. Because relationship tags linked characters together, any change in one kind of standardization—for example, how nicknames were handled—needed to be synchronous with other necessary changes to avoid successive rounds of tag revisions which required significant wrangler labor and were particularly taxing on TFR's server:

> But, so then it's like, do we go through, yes we definitely have to go through all of Ned's [Ned Stark's relationship tags] and fix all of the canonical relationship tags. That's going to be an entire day's worth of project probably, to make sure everything is as it should be. And hopefully none of those then have to be revised again because somebody else in the relationship also got a nickname added, and, oh my god. (Justine, Interview)

Here, Justine described three sources of inertia preventing the changes she saw as necessary to the existing tags in her fandom. First, despite her certainty that most names in her fandom were amenable to standardization, prominent exceptions such as "Cersei Lannister" made it difficult to articulate a single rule for alternate names. Second, changes to each character tag required changes among dozens of linked relationship tags. Third, relationship tags necessarily linked more than one character tag, requiring coordination in the timing of character tag changes. Like Victoria, Matilde, Brigid, and Amita, Justine noted that decisions on character tags were complex because of elements inherent to the source material and users' fanworks: in her case, that source material creators name characters differently to reflect different attitudes towards marriage and fanwork creators combine characters in numerous relationship combinations.

145

The examples in this section detailed relatively mundane and everyday issues around names—adoption, marriage, and nicknames. For each single character, fans, users, and wranglers had no confusion as to the relationship between the tag term and the referent. However, contemporary and ongoing source material necessitated ongoing updates not captured by guidelines, so that previously simple character-name relationships became complicated and wrangling decisions were retroactively incorrect.

SUMMARY

In contrast to thinking of names as inherently easier to wrangle than concepts, I show in this chapter that names proved to be active points of contact among many, distinct identities and names proved to change for a number of reasons. While the right name for a term like "Horror" will be relatively fixed for a long period of time—say, for around 200 years of literature—individuals' names change from the course of one book to its sequel or from childhood to adulthood. In trying to decide on canonical character tags, wranglers recognized that names were often shared and dynamic, but had to pick some term that would be stable and understandable to future users.

In the examples I discussed, wranglers described the process of creating canonical names for characters as complex for a variety of reasons. Some of this complexity arose from characters with complex identities. Some of this complexity arose from common phenomena of name changes. In all of these examples, wranglers reflected on what was conceptually correct—which terms and term relationships were authentic to the source material—and what the effect would be for users—such as filtering precision or server speed. For example, they navigated trade-offs between what is authentic in terms of identity and what is effective for filtering given the function of the metatagging at TFR.

146

They also reflected on what is true now and what could be true later, in terms of source material and in user choices.

Where this chapter primarily dealt with character tags and how wranglers made decisions on how to represent names in TFR, the next chapter deals with fandom tags and how wranglers made decisions on how to represent fandoms in TFR. Even more so than character tags, fandom tags were a central organizing facet for the TFR collection, and wranglers carefully considered how best to handle these tags. The instances in the next chapter continue to illustrate the extent to which conflicting classification principles played a role in design decisions central to the functioning of TFR's classification system. Similarly to the above instances of complex decisions on names, the following chapter highlights the embeddedness of wranglers' design within the social and technical infrastructures of TFR.

What's in a Fandom?

In this chapter, I detail three examples in which wranglers faced difficulty and differing opinions in defining fandoms. I use the wranglers' terminology throughout, particularly their sense of the term "fandom." The fanwork community and my participants understood the term "fandom" to refer to a source material work and the activity around it. In cases such as the Bandom example in this chapter, however, a fandom referred to a set of people about which fans generated fanworks. For example, "French History Real Person Fiction" would be a fandom in which the referent was a set of people rather than a source material work. Regardless of this difference, wranglers addressed such fandoms according to the more common fandom type in which fandoms referred to a referent source material work.

As with the previous chapter on names, my focus on fandom in this chapter reveals how wranglers addressed complex classification design decisions ubiquitous across fandoms and central to the operation of TFR. Whereas I framed the design task of deciding on a proper name as an intuitively straightforward task, in the case of fandoms classification theory has provided indications that fandom decisions would be complex. The typical fandom tags referred directly to source material works and the definition of a "work" and relationships among versions, adaptations, and physical instantiations has been a perennial issue for information science (Levinson, 1980; Smiraglia, 2001; Tillett, 2005). The core difficulty is this: intuitively, we might feel that *Romeo & Juliet* has a simple referent, but the multiplicity of Shakespeare's folios complicates this referent, as does the play's existence as a text and as a performance. In the previous chapter, I described how a seemingly stable, unique referent such as a name was more complex than we would expect. In this chapter, I describe how wranglers found works—the

central element of collections from library catalogs to repositories of user-generated stories—to be similarly complex and difficult to resolve in a classification system.

For wranglers, fandoms in TFR were central and complex. Source material works were central to the organization of TFR, such that fandom tags were the user's default browsing tag in the site's interface. For wranglers, fandoms were also the primary method of organizing new user tags for review. Fandoms at TFR were particularly complex because the fanfiction community's hyper-awareness of adaptation, derivation, and extension further blurred the boundaries of "works" (Jenkins, 2013). TFR fandom tags reflected the complexity of relationships among source material works. For example, *Romeo & Juliet* was represented on the TFR as 12 different fandoms, each referring to a different source material work. In addition, the fan community's sensitivity to issues of adaptation and derivation meant that fandom tags also reflected complexity not attended to in other systems: for example, that characters in books wrote books, that cartoon characters watched movies, and that some fanworks themselves became source material for further creation.

The examples in this chapter illustrate two types of decisions wranglers make with regards to fandoms: defining fandom tags and defining the relationships among fandom tags. The first type of decision (illustrated by the American Folklore and Bandom examples) involved issues of terminology common to all decisions in a classification system where the fuzzy boundaries of work in the fanwork context added further complications. These two examples highlight the relevance of the system's embeddedness in the social world of fanworks. The second type of decision (illustrated by the Bandom and Fiction-within-Fiction examples) is one familiar to classification designers in traditional settings, but it was particularly convoluted for TFR given the centrality of adaptation and transformation to the fanfiction community. In particular, the

final example continues the thread from the previous findings chapter in which design decisions were embedded within the technical infrastructure of TFR. Together, the three examples illustrate the extent to which wranglers used creativity to address complexities beyond what they had previously instantiated in the guidelines.

AMERICAN FOLKLORE

In July 2013, 33 wranglers opted into a conclave to discuss topics related to managing religion and folklore fandoms and topics in TFR. While a body of religious stories is a "work" in a different way than a novel or a television series is a "work," the structure of TFR treats them the same: as fandoms that users can designate as the source material or inspiration for their fan works. The religion and folklore (R&L) discussion was structured as a series of seven pre-selected topics, including Character Placement (63 messages) and Fandom-Level Terminology (139 messages). The latter topic was particularly urgent given persistent questions among the assembled wranglers on the use of terms such as "religion," "mythology," or "lore" in fandom tags. For which fandoms in the repository was each term appropriate? Could choosing a single term for all relevant cases ameliorate perceptions of the privilege of European, literature-based traditions over aboriginal, oral traditions? Wranglers participating in this discussion drew from various sources of evidence—warrants—to develop consensus on these questions and create new policies for naming religion and folklore fandoms. The wranglers began this discussion with the shared understanding that they should treat works of folklore with the same respect as they would for works of religion. In other words, ethical warrant or the priority of creating an inclusive and respectful classification system was the instigating factor for the discussion that ensued. Following the wranglers' framing of these issues, I do not distinguish between "religion" and "folklore" in my analysis of this example.

In introducing the Fandom-Level Terminology conversation, Amy, the staff member leading the discussion included the following question: "Is NATION Folklore a useful division, for users and wranglers? Or is there a better way to divide folklore systems?" Responding to this question, a wrangler named Sue offered a specific fandom in need of revision:

> American Folklore - As has been pointed out, the way in which this currently is being used is a bit of a misnomer. Could we perhaps consider changing this to something like United States Folklore in order to avoid it being thought that we're being Americentric or that this should cover the lore of various other parts of the Americas? (Sue, R&L Discussion)

Though Sue presented a number of fandoms with similar problems, the wranglers participating in the conclave quickly picked up "American Folklore" as a key example. Thirteen other wranglers contributed 38 messages and developed a set of suggestions and concerns around the fandom and its referent. This discussion around "American Folklore" engaged with two interrelated decisions: what is the referent of this fandom and what should be the term for this fandom?

With regards to the first question, wranglers noted that the term "American Folklore" was not easily linked to a single religion, in part because American folklore is composed of elements from immigrant traditions and associated with peoples and cultures not easily separated by the northern or southern borders of the United States of America. Adding to this ambiguity was the term "American," which was a perennial issue on TFR as it was often used to mean "originating from the United States of America" though readers could have interpreted it as having the broader meaning, "originating from North America and South America." Some wranglers also noted that American folklore stories were entangled with stories about known historical figures such

151

as Johnny Appleseed, adding further ambiguity by crossing the fuzzy boundaries between religion and history.

Before they had fully resolved these questions about the referent, wranglers attempted to solve the second problem: the term for the fandom. An early reply to Sue's comment offered an alternative fandom tag term with the potential to answer both issues, "I think specifically with American stuff, we can label it Tall Tales & Folklore instead of 'lore'" (Julian, R&L Discussion). The proposed term generated controversy and became the focus of the discussion around "American Folklore." "Tall Tales" was a controversial term for two reasons. Pat, an experienced wrangler voiced the first reason:

> Why? Why would we privilege one particular culture's label with a different name? One of the issues [TFR] has had from the beginning is the perception that we're completely US-focused, and that our active interest outside that is...token, if not lip service. While the truth of such perceptions is debatable, we need to be aware of it. (Pat, R&L Discussion)

Follow-up questions and responses from other wranglers clarified that Pat was concerned about using a local term that had ambiguous meaning outside of the American context. In other words, "Tall Tales" had a specific meaning for English speakers raised in North America, but had the potential to be understood by other English speakers as a generic term for fanciful tales, having no particular reference to American folklore. Another wrangler, Rachel, expressed this objection in another way:

> Agreed that we should prioritizie [sic] insider names wherever possible (which isn't going to be always; sometimes the [TFR] standard will be the best we have), but we should also consider what the names we choose would seem like to persons not privvy [sic] to insider wrangler knowledge. (Rachel, R&L Discussion)

For sacred traditions, which are related to personal identity and which can be the subject of discrimination, wranglers were particularly concerned that they practice respect for insider or local terminology. Where using local terminology created ambiguity, wranglers

typically used the fandom name itself as a disambiguation in parentheses. For example, "Nanna (Norse Mythology)" was a wrangler-created term that resolved the ambiguity of the name "Nanna," a Norse goddess, by adding the fandom name for Norse mythology in parentheses. In the case of a fandom tag, the impact of choosing a clear and respectful term was amplified by its likely use as disambiguation for terms under its umbrella, and because the term had to stand on its own. That is, "Norse Mythology" had to be unambiguous and respectful, as wranglers would use it as a parenthetical disambiguation on dozens, if not hundreds, of tags.

The question of the referent of "American Folklore" returned during the debate over the term "Tall Tales," leading to the second reason wranglers found the latter term controversial. Some wranglers, such as Mackenzie, objected to "Tall Tales" as a preferred term for "American Folklore" because they perceived an incongruity between the two terms' referents:

> "Tall Tales" is a term used for a *certain subset* of US folktales - the ones around Pecos Bill and Paul Bunyan and the Celebrated Jumping Frog and, like, Bowie. It wouldn't apply to things like Appalachian folklore or Br'er Rabbit or New England ghost tales or Caddy, though. So while I support using these when we can, we need to be careful about using them properly (especially from places other than the US, when we won't necessarily have anyone familiar with shades of meaning). We'd probably still want some kind of overarching metatag to include them all, not just the ones the specific term includes. (Mackenzie, R&L Discussion)

As Mackenzie explained, "Tall Tales" was both too narrow to be inclusive of the folklore of the United States of America, as there are other folklore traditions within the country, and too broad to authentically represent US folklore, as some tall tales crossed national borders, as more than one wrangler pointed out was the case for Paul Bunyan. This objection was distinct from Rachel's comment about how the term "Tall Tales" might be

perceived beyond North America, and illustrates further the complexity of the referent question.

The ambiguity of the term "American Folklore" drove many of these problems. Any suggested term that was more concrete in definition would either have mismatched the "American Folklore" referent because the term would have been too specific, as in the case of "Tall Tales," or have been so broad as to have had a thin connection to the referent, as in the case of "Lore of the Americas." In the latter case, wranglers noted that collecting all legends from North and South America together would accomplish neatness but little else. Wranglers who objected to such a broad category argued that there was such a variety of religions, traditions, and cultures within that geographical area that referring to them as a single body of works was inauthentic. In response to the suggestion to use "Lore of the Americas," Amy stated:

> This is the problem with using geography: it erases deeply felt cultural differences within a geographical area, as well as deeply felt cultural connections outside that area.
>
> So anyway, this issue is broader than the U.S. alone. [Mackenzie] touched on the problem of dividing up the British Isles by geography, and we've already decided that "Middle East" is an unacceptable term/grouping for the archive. "East Asia" would also be unacceptable, to my mind.
>
> I'm personally happier to break fandoms up into smaller categories where needed, and I'm inherently suspicious of metatagging in this context unless it's been checked for both inoffensiveness and usefulness for people who participate in the fandom. (Amy, R&L Discussion)

Here, Amy expressed concern that broad, geographic terms would provide structure without meaning. Her comment shows that simple or tidy solutions were not necessarily good ones; dividing the world up into regions was possible but would not serve the users in terms of respect for individual cultures or in terms of filtering.

For the handful of individuals who created fanworks in the "American Folklore" fandom, the meaning of the term "American Folklore" may have been more instinctually recognizable than could be captured by an unambiguous term that would direct unfamiliar users to a stable referent. In other words, the users who tagged works as "American Folklore" obviously felt the fandom name had meaning, even if wranglers could not define what that meaning was. The "work" in this case was not a fixed set of stories or legends, as might be the case for "דומלת | Talmud," but a looser set of stories wranglers could not, with confidence, identify or name. Even viewing "American Folklore" as a superwork would not resolve this ambiguity, as wranglers would still need to decide which set of fandoms would belong to it as subtags.

The complex issue of "American Folklore" did not take place in isolation; the design decisions in this system were interdependent and ongoing changes to the system had to be considered. Specifically, the wranglers agreed to postpone many of the proposed changes in the R&L discussion given an anticipated change in the metadata structure of TFR: a new metadata attribute for fandom tags, specifying country (or countries) of origin. Referring to fandom tags that use country names, Amy went on to explain,

> Especially with [country of origin] metadata on the horizon, these tags are redundant, and they both erase distinctions between different story cycles in the same region and enforce some sort of connection between completely unrelated works/characters. (Amy, R&L Discussion)

As Amy explained, fandoms such as "American Folklore," with referents and terms related to country-level geography, would have to be reconsidered again after the anticipated metadata change. Wranglers discussed whether the metadata change might resolve the underlying ambiguity of "American Folklore," making their proposed solutions redundant. The other possibility—that the metadata change would add different

155

problems—also generated support for the decision to delay changes to the "American Folklore" fandom.

Given the ambiguity of the referent and the objections to alternate terms, it is not surprising "American Folklore" was not among the fandom names for which R&L discussion participants reached a consensus solution. While the fandom's wrangler and her peers agreed that "American Folklore" was a flawed term with recognizable problems, the conclusion was that the term could not be changed because of its intersections with other "wicked" problems (Buchanan, 1992) within TFR, including the as-yet-undefined relationship between fandom tags and countries of origin, the inconsistency of respect given to supernatural traditions associated with more or less privileged cultures and languages, the particular role of America as an immigrant country, and "America" as having a strategically-deployed ambiguous meaning.

BANDOM

Similar to "American Folklore," the fandom term "Bandom" presented wranglers with an unresolved discrepancy between referent and terminology. Whereas "American Folklore" had geographic, cultural, and religious issues, "Bandom" was part of an ongoing controversy among fan communities. To understand the difficulty that wranglers had in managing this term within TFR, I first review the controversy from the broader context of music fandom by synthesizing information from two sources: statements from participant interviews and the fan community resources wranglers referenced in their efforts to explain the nature of the controversy. Second, using system data, I describe the state of usage of the "Bandom" fandom tag at TFR. Third, and finally, I draw upon participant interviews and system data to explore how wranglers managed the "Bandom" fandom tag at TFR.

The history of "Bandom"

The term "Bandom," simply, was a portmanteau of "band" and "fandom." This simplicity invited controversy, as fans disagreed over which bands—and therefore, which fans—the term included. As one wrangler explained,

> When the term started coming into place two groups of people used this, one being Fall Out Boy and related fandoms like MCR [My Chemical Romance] and Panic! at the Disco and another one being a lot of the rock music fandoms. (Dani, Interview)

Fans of a set of interrelated emo rock bands—Fall Out Boy, My Chemical Romance, and Panic! at the Disco—adopted a term that had been in use for decades by rock music fans. The convergence was not surprising; the portmanteau was simple and provided no signal of exclusivity. In other words, unfamiliar users thought of the generic term, once it was used in systems such as TFR's, as being applicable to their particular band fandom. Fans of more bands came to use the term, adding to its older, classic rock usage and its usage in the recent and prolific Fall Out Boy following. A second wrangler described how these multiple usages confused her, referencing her experience as a member of one fan community and as a wrangler:

> I was in Bandom very heavily in the sort of Fall Out Boy, 2000, what was that 2005 or 2006, you know, to 2010 period. So it meant something very specific to me. And when I volunteered to take it on [as a wrangler] I was like, "Cool, I'll take that on," and then sort of turned up and then, "by the way, this is a fandom, this a term that some would say stolen, some would [say] borrowed, from 80's fandoms, so they have to be allowed to stay here too." And then as time has gone on, there's a new generation of fans, I think, sort of, picking up that banner now. So suddenly I have a whole bunch of new bands showing up in the Bandom tag that I've never heard of. (Kerri, Interview)

From Kerri's description, it might seem as though no single group of fans was aware of the term's usage in other communities, at least not until they looked closely at a terminology system such as TFR's. However, the debate around the term had a history as

157

long as TFR itself and occurred on multiple platforms, most notably LiveJournal, a blogging platform popular among the fanfiction community. Fanlore.org, a wiki of fan community topics, singled out November 2007 and April 2008 as being the periods of intense debate among influential LiveJournal bloggers during which fans discussed phenomena of co-opting, exclusivity, and prescriptive use. LiveJournal participants characterized the discussion as "distancing" and full of "vitriol" ("Bandom Terminology Debate," n.d.).

During the LiveJournal debates, fans suggested alternative terms to designate particular subsets of band fandom such as "eFandom" for emo bands such as Fall Out Boy and "Pete Wentz & His Humans" for bands connected by Fall Out Boy's bassist ("Bandom Terminology Debate," n.d.). The appearance that outsiders were prescribing usage caused problems for both proposed solutions. Specifically, the Fall Out Boy community rejected "eFandom" because the individuals who proposed the term did not belong to the fan community to which it referred and imposed the term without consultation—including by defining the term in general-use fan resources ("Bandom Terminology Debate," n.d.). Similarly, Dani explained that while the term "Pete Wentz & His Humans" served the function of clearly designating the related bands at the heart of the "Bandom" debate, the community rejected it:

> "Pete Wentz & His Humans" was another name trying to get just that one half covered and it didn't work out because fans decided that that wasn't good enough… "Pete Wentz & His Humans" was an attempt to do that [give a common name to related bands] for some of the fans but the fans didn't like that they couldn't use Bandom because that was their name. (Dani, Interview)

Dani noted that the terminology debate, in part, remained unresolved because multiple groups had become attached to "Bandom" as a name for their fandom and resisted alternatives as attempts to make them relinquish their claim. The LiveJournal debates of

2007 and 2008 ended with "Bandom" remaining an ambiguous term and participants agreeing only to stop attempting to prescribe usage across fan communities ("Bandom Terminology Debate," n.d.).

That the LiveJournal debates left "Bandom" as an ambiguous term is evident in the usage of the term at TFR. In the following subsection, I describe the extent to which "Bandom" had come to refer to any band fandom, becoming larger in usage than the generic tag for musician fandom at TFR. I first describe the coverage of "Bandom" at TFR in terms of hierarchical structure and then in terms of overlap of usage.

"Bandom" at TFR

On TFR, "Bandom" was a popular fandom tag largely interchangeable with the "Music RPF" (Real Person Fiction) tag. Wranglers placed it within "Celebrities & Real People" and "Music & Bands," top-level categories—not tags—meant to organize fandom tags by media type. The placement of "Bandom" was parallel to the fandom tag "Music RPF," though in the common-sense definition and in wranglers' discussions, "Bandom" was a subset of "Music RPF" fandom. In other words, while "Music RPF" would refer to works about any musicians, "Bandom" would refer only to bands and not to solo musicians or classical composers, thereby being a subset of "Music RPF." Wranglers designated fandom tags as direct subtags of either or both tags or as subtags to more specific genres below "Music RPF." For example, "Wolfgang Amadeus Mozart" is a subtag of "Classical Music RPF," which is a subtag of "Music RPF." "Pink Floyd" is a subtag of both "Bandom" and "Music RPF," and "Twenty One Pilots" is a subtag only of "Bandom." "Bandom" therefore functioned as a superwork, collecting related works that shared the quality of being Bandom fandoms.

Because "Bandom" served both as a subset and as a parallel term to "Music RPF," there was an imperfect overlap between the "Music RPF" and "Bandom" tags, as shown in Figure 14. As of March 2017, a search for "Music RPF" returned 33,948 fanworks, with very few (739) tagged with "Music RPF" itself. In the vast majority of cases, users had tagged with more specific subtags (e.g., "Wolfgang Amadeus Mozart") and not used the "Music RPF" tag itself. The usage of Bandom was much different. A search for "Bandom" returned 36,641 fanworks. For 13,258 of those fanworks (more than one-third), users had tagged for "Bandom" directly. The usage of Bandom suggests a different meaning on the part of users, as users appeared to treat "Music RPF" as a category and "Bandom" as a fandom in its own right—the first organized a set of fandoms but was not used directly while the later was frequently used directly. "Music RPF" and "Bandom," through users' tagging with overlapping subtags or both fandom tags, shared 25,699 works, or approximately 60% of the total unique works across both sets. Figure 14 illustrates this substantial overlap between the usage of the two tag hierarchies. I generated Figure 14 from the work counts of each metatag using a Venn diagram generator (Bioinformatics & Research Computing at Whitehead Institute, 2009).

"Music RPF" alone	8249	
"Bandom" alone	10942	
Overlap	25699	
Total	44890	
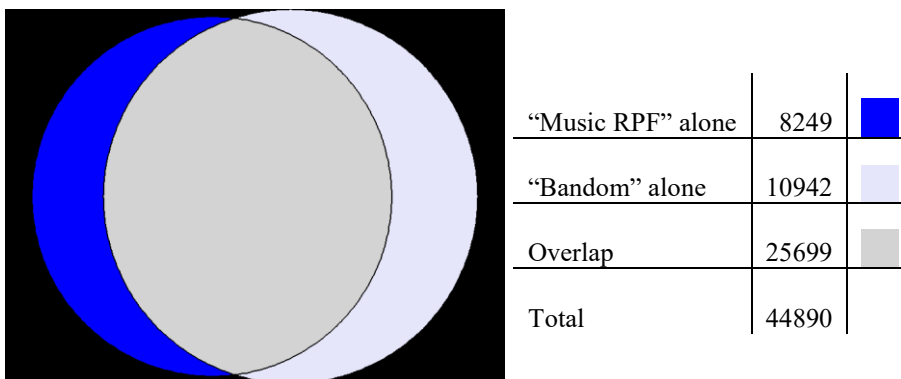
Figure 14 Proportion of works belonging to "Music RPF" and "Bandom"

My Chemical Romance, one of the bands central to the 2007 and 2008 Bandom terminology debate, accounted for 12,926 fanworks in (or, approximately 50% of) this overlap. The 8,249 fanworks among "Music RPF" exclusive of "Bandom" included works based on solo musicians such as Adam Lambert and Taylor Swift and some band fandom tags wranglers had not added to the "Bandom" tag, such as "One Direction," "Daft Punk," and "Muse." Users primarily tagged the 10942 works exclusive to "Bandom"—without "Music RPF"—with a few bands that wranglers had not yet added to "Music RPF," such as "Twenty One Pilots" and "Tokio Hotel."

The effect of the state of usage was that the "Bandom" tag was so large and general that it provided little filtering capability for users. The number of works within "Bandom" was so large that users were unlikely to keep up with reading the entire fandom; more likely, users were interested in a subset of the bands and filtered according to those individual fandoms. The following section, in which I unpack the wranglers' perspective, explains how the "Bandom" fandom tag came to be this way and how wranglers viewed the impact on users.

"Bandom" wrangling at TFR

"Bandom" was already a controversial term by the time TFR launched. Among TFR wranglers, an early solution to the "Bandom" controversy was to recategorize "Bandom" as a freeform tag rather than a fandom tag, which allowed wranglers to dissociate it from any specific set of bands. As a freeform tag, "Bandom" would have been a generic concept, similar to "Horror," applicable to any fandom. Users could filter by the "Bandom" freeform tag and return all those works users had tagged as "Bandom." Despite the utility of this solution, within months, users had recreated "Bandom" as a

fandom tag, leaving wranglers to decide how to relate it to other fandom tags. Summarizing wrangler staff discussion in 2009, one wrangler stated:

> We really needed to take the [wrangler] team out of the debate. […] As was discussed before: Either Bandom means only a few bands, or ALL bands. […] This is a no-win situation for us. So, we're taking ourselves out of the war-of-the-definitions. (Staff meeting minutes, Internal wiki)

The policy that emerged from wrangler discussion in 2009 was for association with the Bandom metatag to be determined by use, with no prescriptive boundaries set for wranglers. As a wrangler stated on the Bandom tag in 2010, "To clarify: in general, if a band's fandom considers itself Bandom, it gets added under the metatag" (internal system data). The internal wiki page for "Music RPF" wranglers provided similar instructions: "Users may tag works containing any artist as bandom. Wranglers will not dictate what does and does not belong in any genre, including this one" (internal wiki). In practice, this meant that if a user tagged a work with the fandoms "Coldplay" and "Bandom," wranglers would add "Coldplay" as a subtag to "Bandom." This policy of ambiguous use centered user tagging choices as defining the superwork fandom tag, "Bandom."

As the numbers in the previous subsection illustrate, searches for Bandom on TFR returned fanworks using a wide variety of bands as source material, though the majority of the fanworks under this metatag were about Fall Out Boy, My Chemical Romance, and Panic! at the Disco. No metatag existed that collected together only this set of related bands. Wranglers, as users, experienced the usage as a problem:

> I was being frustrated by this as [a TFR] user quite recently - if I were looking for Hanson AUs [Alternate Universes], I'd go directly to them, not look for them under Bandom; if I'm looking for Bandom AUs, I'm NOT looking for Hanson. So I've got to say that I doubt that Bandom as it's being used as a meta-tag right now is doing much of anybody much good. I mean, I define Bandom FAR more broadly than most people involved in the Six-Degrees-of-Pete-Wentz-and-MCR world do […] and using it to filter the way the current filtering system is set up does me far less good than it ought to. I expect it's a world of frustration for

people who are into five bands out of that list rather than 25 (TOTALLY making up numbers here). (Internal system comment)

This wrangler, who posted this message to the "Bandom" tag, illustrated the negative effect on filtering the ambiguous use policy created, as well as the strong emotional reaction this negative effect caused.

The frustration the wrangler above expressed is understandable given the close association among this subset of bands. The lack of a clear and exclusive term belied the strong associations fans made among this subset of bands. To demonstrate the complexity of associations among the bands, Fanlore.org's article ("Bandom (Decaydance+, My Chemical Romance)," n.d.) features a fan-developed flowchart (Figure 15). I provide this chart not to explain the associations between the bands but as evidence of users' perception of the bands as having many—and many kinds of—associations.
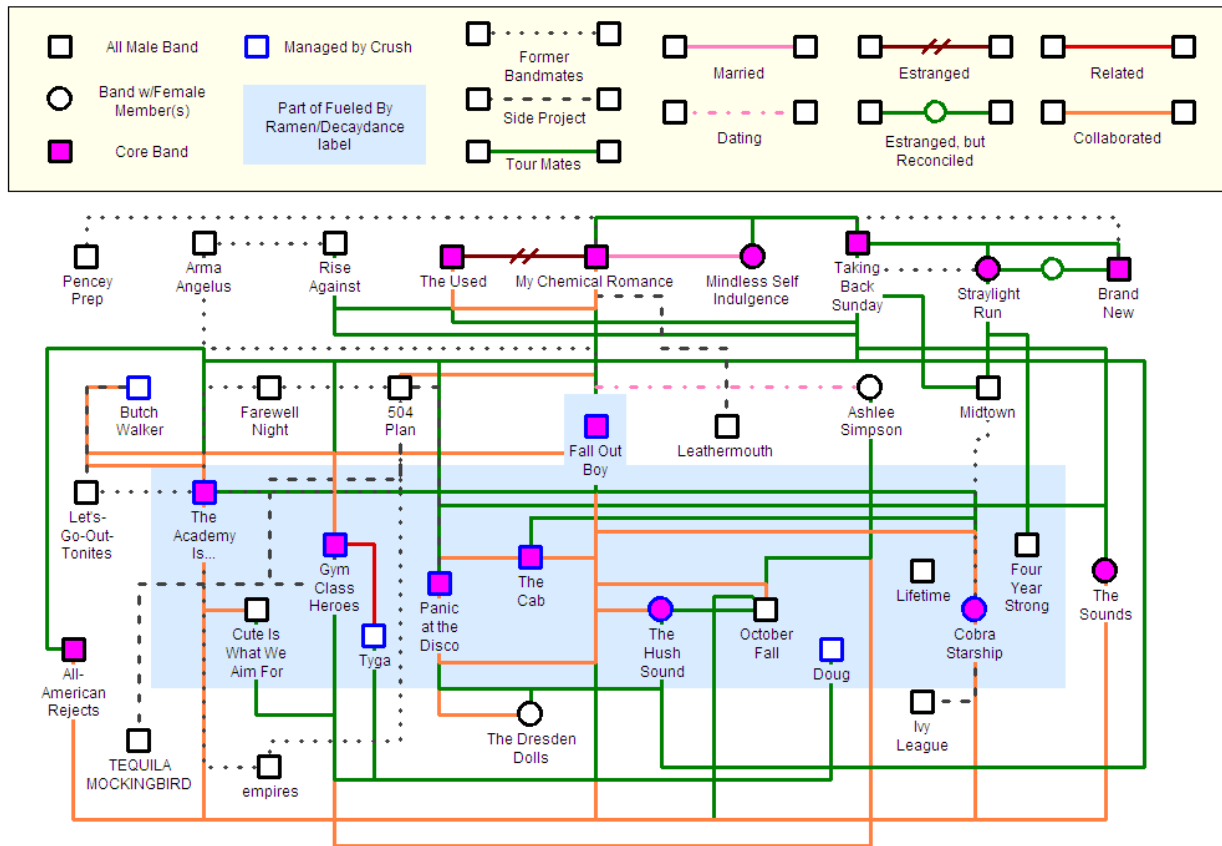
Figure 15 Flowchart appearing on Fanlorg.org's Bandom (Decaydance+, My Chemical Romance) article

The overlapping associations among this subset of bands—which Fanlore.org refers to with the unwieldy term "Bandom (Decaydance+, My Chemical Romance)"—contributed to a sense of unified fandom. The heterogeneity of kinds of associations—collaborations, romances, tours—made it impossible to identify and name a single shared characteristic. Similar to the users whose choices were relevant to the "American Folklore" example, users in this case may have felt there was a coherent referent to "Bandom," but wranglers could not identify or name this referent.

In practice, fanfiction creators commonly wrote across many of the associated bands at once, so that the network of characters in fanworks would resemble the tangle

164

illustrated in Figure 15. This community of creators was particularly prolific; in 2016, users tagged an average of 157 new completed works per week as "Bandom," with 43 works or 27% of those tagged with "My Chemical Romance" and 23% tagged with "Pete Wentz." For TFR readers interested in keeping up on fanworks belonging to that subset, neither the "Bandom" metatag nor an individual subtag like "My Chemical Romance" or "Pete Wentz" returned this set with reasonable precision and recall. That is, the "Bandom" tag was not precise, and returned many irrelevant works, and the individual subtags did not have complete recall, missing many relevant works tagged with other subtags. The operation of the TFR tag search system was such that searching for works on any of these bands requires a metatag; the system supported only the Boolean AND operator within the "Fandoms" search field for works. Therefore, if a user wanted to see works tagged with "My Chemical Romance" OR "Panic! at the Disco" OR "Fall Out Boy," he or she had to perform three different searches. This user would likely receive many of the same results from each of these three searches, as fans commonly wrote about more than one of these three bands within a single work. As a result, the user would likely spend additional time identifying the search results that were unique to each of the searches, if they did not wish to risk missing one.

It is unlikely that this discrepancy between terminology and referent would ever be resolved, given the reasons explored above: that the fan community had become attached to an ambiguous term and that the heterogeneity of associations among the bands defied an obvious alternative. Within the context of TFR, there was the additional problem that the agents who would enact a solution, the wranglers, were themselves a part of the communities at the heart of the controversy. As Dani explained,

> There's this divide between who is and who is not Bandom, so you have to get
> people who are impartial to it on that tag. It's very difficult for us to all do a

megafandom talking to each other because it's got that divide and we're not impartial because we're obviously deeply invested in our fandoms. (Dani, Interview)

Wranglers typically took on responsibility for a fandom because of their familiarity with the source material, their familiarity with its fanworks, and their interest in supporting fan activity on TFR. Wranglers were not disinterested administrators of the system but were embedded within the collection and its controversies. As representatives of the communities inconvenienced by this terminology problem, wranglers were motivated to find solutions but resistant to consider actions that their communities had previously rejected. Relationships among fandom tags put wranglers in the difficult position of negotiating across contentious fan community divides.

"Bandom" remained an ambiguous tag and TFR wranglers yielded the meaning of the term to users. Despite deeply held feelings by the wranglers themselves and their insight regarding the negative effect for filtering, they determined that any intervention beyond the current policy would exclude some users.

These first two examples of "American Folklore" and "Bandom" illustrate cases in which TFR wranglers had to engage with difficult and long-lived controversies with no clear solution. The phenomena that prompted the difficulties had histories beyond TFR itself and wranglers could not solve them within TFR. The next and final example illustrates a case in which a phenomenon originating within TFR—a creative and novel fanwork event—spurred thought-provoking complications regarding fandom tags, works as referents for fandom tags, and the relationships among them. Though the phenomenon in the "Fiction-within-Fiction" case that follows did not have a controversial history beyond TFR, the wranglers who made decisions on fandom tags weighed similar considerations as in the cases above. In combination with the first two cases, this final case illustrates that, even when the wrangling team had a great deal of autonomy in

creating definitions and boundaries, persistent conflicts among the factors of authenticity, filtering, and user choices arose.

FICTION-WITHIN-FICTION

In January 2014, TFR hosted a community event, called "Fiction-within-Fiction," which was a "ficathon" in which users could nominate source works which other users could then choose from in writing stories for the event. The theme of this ficathon was works that existed within fictional works, such as the 1950's television series *Pleasantville* that provided the setting for the 1998 movie *Pleasantville* or *The Itchy and Scratchy Show*, a television series that the children on *The Simpsons* watched religiously. In hosting the event, TFR hosted the stories and facilitated the creation of a set of tags that event moderators used to coordinate the event. In the case of the Fiction-within-Fiction event, this meant that nominated works such as *The Itchy and Scratchy Show* could appear as fandom tags in the set organized by their media type, in this case, "Cartoons."

The wranglers discussed this event in detail. Part of the purpose of event tag sets was to give the wranglers an opportunity to manage these tags before the majority of participants added their stories or artwork to the site. Since events involved users posting and reading many related fanworks in a relatively short span of time, it was particularly urgent for wranglers to integrate these new tags into the curated folksonomy. Managing the tag sets for events such as Fiction-in-Fiction was a different type of work and challenge for the wranglers than their typical work at TFR. Maintenance and reaction were the dominant modes of work at TFR: wranglers could not designate tags as canonical tags or place them in relationships with other tags until a user created them—or something similar—as a tag. Managing tag sets, on the other hand, was instead a matter

167

of planning and creating: in the early stages of an event, wranglers had the chance to make decisions about tags and their relationships before the tags were "live" in the system. This mode of work was especially characteristic of cases like the Fiction-within-Fiction event, which, given its subject matter, targeted fandoms that users had not yet added to the collection. For wranglers, ficathon events provided an uncommon opportunity to develop guidelines and apply them consistently from the outset of a trend. In the cases of the few Fiction-within-Fiction fandoms that existed in the site before this event, wranglers had made ad hoc interpretations of the existing guidelines to handle these atypical cases. During the early stages of the Fiction-within-Fiction event, wranglers discussed possible guidelines based on these few existing fandoms and the wide array of nominated fandoms. As with the typical maintenance-and-reaction work, the deliberation around Fiction-within-Fiction was about finding a solution that fit the currently known tags and that wranglers could reasonably extrapolate to cover expected or likely future tags.

Wranglers discussed their preparation for this fandom event over the wrangling mailing list, the chatroom, and in system data. Toni, a wrangler responsible for managing the event's tag set, started the mailing list discussion by asking for the group's opinion on four questions:

- In which media type (Cartoons, Books, Other Media, etc.) should the [Fiction-within-Fiction] fandoms be categorized, considering that some differ in media type from their source fandoms?

- If they need disambiguation, do we do so as if they were a regular fandom, by adding their media type, by adding the title of their source fandom, or by adding "fictional," "imaginary," or another chosen term?

- Should they be automatically subtagged to their source fandoms if they are made canonical?

- Should they, at any point, be made synonymous to their source fandoms?

These four questions touched on four different decisions wranglers made about fandom tags. The second, third, and fourth question are relevant to understanding complications in classification design decisions around works, while the first question is beyond the scope of this topic. The third and fourth question were opposing alternatives; since wranglers came to a quick consensus to make Fiction-within-Fiction fandoms canonical rather than synonyms, I do not include the fourth question in this analysis.

In the subsection "Problem cases," I explore how wranglers invoked problem cases to object to or support proposed disambiguation solutions as they addressed the second question. This exploration illustrates how the variety of fandoms users nominated for this event created complexity for the wranglers, and how wranglers navigated this complexity by addressing or scoping out problematic fandoms. In the section "Authentic vs. effective classification," I explore how wranglers argued for or against subtagging and other hierarchical relationships based on the authenticity of the relationships or their effect on filtering. In doing so, wranglers put into conflict two alternative purposes of classification systems—whether the classification should authenticity represent an external reality (as argued by scholars such as Bliss and Hjørland) or whether classification systems are more importantly a means for retrieving items (as argued by scholars such as López-Huertas and Bates). By probing wranglers' arguments with respect to Toni's third question, I show how wranglers understood the classification structure of the hierarchy as representing authentic relationships or aiding filtering, and how they addressed conflicts between these two purposes. In sum, I chose these two questions to explore among the four Toni put forth because they represent the two types of classification design decisions around fandoms: what is a work, and what are the relationships among works?

Problem cases

Wranglers used problem cases—especially difficult examples of fandoms relevant to the event—to think through proposed solutions. Throughout the discussion on the Fiction-within-Fiction event, wranglers used problem cases to challenge or occasionally to support particular disambiguation strategies. For example, a wrangler used *The Princess Bride*[5] as a challenge to one proposed solution. The wrangler who first used this novel as an example argued against disambiguating Fiction-within-Fiction fandoms by adding a fandom's source work as a parenthetical, since "The Princess Bride (The Princess Bride)" would likely confuse readers. Alternatively, wranglers used *The Princess Bride* to support a different solution: to treat these fandoms as if they were any other fandom. Wranglers who supported this approach explained that it would be easy for fans familiar with the book to distinguish the fandom tag "The Princess Bride – William Goldman" from the fandom tag for its imagined source material, "The Princess Bride – Simon Morgenstern." In this manner, the trickier problem cases, once solved, became strong arguments for the solutions that satisfied them.

Parallel to the discussion of the above solution, to disambiguate the Fiction-within-Fiction fandoms using the same elements (such as author) used in conventional fandoms, the wranglers discussed Toni's suggestion to use "(Fictional)" as a disambiguation. Several wranglers supported the idea of adding "(Fictional)" as a disambiguation in the fandom tags when necessary, such as to distinguish a Fiction-within-Fiction fandom from a source fandom of the same name (e.g., "The Princess Bride (Fictional)" or "Pleasantville (Fictional)"). One wrangler, Jess, objected to this solution, referencing one fandom a user had nominated for the event:

---

[5] *The Princess Bride* by William Goldman is a novel the author presents as the abridged, annotated version of Simon Morgenstern's novel, *The Princess Bride.* Neither Simon Morgenstern nor his novel exist, but William Goldman and his novel *The Princess Bride* do.

I'm going to object to the idea of the disambigs [disambiguations] including the word fictional [or] imaginary for two reasons. (1) why is something any more fictional or imaginary just because it has been sourced out of a work that was fiction itself? (2) I have seen some of the nominations rolling in and they include that ever-present sticky-wicket [difficult circumstance] on [TFR] of religion and lore and identity fandoms. The Gospel Of Eve has been nominated. It is lost source text apocrypha from the bible. I think it would be a really bad idea to apply words like imaginary or fiction to a part of anyone's religious or cultural identity just because the source work itself is no longer extant. That's my two cents for the minute. Take it or leave it as you like as always (Jess, Mailing list)

Jess used a specific problem case—"The Gospel of Eve"—to make her point, but she used it to represent a class of fandoms for which the proposed solution would have been misleading if not offensive. Jess was referencing a mutual understanding among the wranglers that any terms which touch on identity—such as religious concepts, sexuality, ethnicity, and gender—were treated with extra care, so that wranglers did not accomplish efficient organization and filtering by risking offense or harm to TFR users.

Like Jess, wranglers throughout the Fiction-within-Fiction tag set discussion invoked problem cases to test the feasibility of proposed solutions, often challenging proposed solutions by showing how they failed to address a particular case. When wranglers used problem cases in this way, they made two implications: that the identified case was within the set of those that needed to be addressed by the solution and that this case was representative of others that might reasonably occur in the future. In the course of answering the question of disambiguation, wranglers challenged the first of these two implications by rejecting the relevance of "The Gospel of Eve" problem case to their deliberations. In the most concise version of this argument, Nica stated in the mailing list discussion, "I think we should remember that not necessarily all fandoms used in this particular ficathon should be under the same guidelines -- just because a group of fans using [TFR] are considering them the same thing, it doesn't mean we need to." Nica and other wranglers in this discussion claimed that cases that defied the logic of every

171

solution were actually beyond the intended scope of the event and therefore beyond the scope of any relevant guideline for disambiguation. Rather than deciding on a term that fit Jess's problem case as well as the more conventional examples, Nica used Jess's problem case an example of how the fandoms users added to the event did not perfectly correspond to the types of works wranglers sought to address. While the Fiction-within-Fiction event spurred the wranglers to discuss how to handle works that exist solely within other works, the users in the event were capable of breaking their own rules and nominating a work such as *The Gospel of Eve*—one known through another work but understood to have an existence beyond it. In making this argument, Nica supported the second implication—that a case such as "The Gospel of Eve" might occur in the future—but advised her peers to recognize such cases as out of scope.

The reaction to the problem case of "The Gospel of Eve" went beyond the cliché of "the exception that proves the rule." Nica did not just dismiss Jess's problem case as being too unique to be addressed by the guidelines. Rather, this example served as a reminder that—regardless of how completely and correctly the wranglers defined and arranged the terminology of the system—users were always capable of misusing it. One of the major limitations of TFR's curated folksonomy design was that wranglers could not alter users' tags, only change how those tags relate to each other. This limitation provided opportunities for users to introduce unresolvable contradictions to an otherwise coherent system. Pat provided a similar comment in response to the case of *The Princess Bride*, claiming apprehension that a mistaken user was going to see the fandom tag that listed Simon Morgenstern as the author and submit a Support ticket to correct it to William Goldman. Pat and Nica, as experienced TFR wranglers, shared the expectation that users could always misunderstand and misuse the classification system. Pat and Nica's claims redirected the conversation away from impossible problem cases by taking

the stance that user error was inevitable and not grounds for rejecting otherwise sound guidelines.

While the wranglers' use of *The Gospel of Eve* and *The Princess Bride* resulted in wranglers agreeing on which works were beyond the scope of the Fiction-within-Fiction event, Toni's questions remained difficult to answer even within the narrower set of relevant works. In the following subsection, I describe how wranglers experienced conflict between two conflicting aims—authenticity in representation and effectiveness for users—in making decisions on disambiguation and hierarchy.

Authentic vs. effective classification

In the course of the discussion around the Fiction-within-Fiction event, wranglers came to argue over two competing interpretations of the purpose of a classification system: to authentically represent works or to facilitate effective user navigation and filtering. As I illustrated in my literature review, even the first interpretation can have multiple versions: to represent the organization of things as they really are in the world or according to the documents' or authors' own logic. This conflict was common in TFR and reflects the long-standing debates in classification theory and design typically represented by appeals to alternate warrants—consensus warrant against literary or user warrant. In the Fiction-within-Fiction event, this conflict complicated two questions: what name to give these types of fandoms, and what relationship to give these fandoms to their source fandoms.

The first conflict between authenticity and effectiveness occurred in addressing the question of disambiguation. Jess's message, above, pointed out two flaws in using "(Fictional)" as the disambiguating term to denote the difference between these Fiction-within-Fiction works and those of conventional works. Wranglers resolved one flaw—

that some nominated fandoms should not be described as fictional—as being out of the scope of the discussion, as illustrated in the previous subsection. However, multiple wranglers reiterated Jess's concern with the other flaw regarding "(Fictional)"; these wranglers argued that "fictional" was a largely meaningless term in the context of fanworks. Even if the wranglers accepted that originating from within another work made a fandom "more fictional," they did not expect users to make the same distinction, especially as users would encounter these fandom terms alone and not in the aggregated view of the wrangler. In other words, users would only see a fandom labelled "(Fictional)" and not see the full set of which fandoms did and did not have this disambiguation, clarifying its meaning. Though it was originally her suggestion, Toni later revealed in the mailing list discussion she hesitated to use "(Fictional)" because users might treat it as a distinction from real-person fiction, stories based on personalities in the real world. Heidi, another wrangler, pointed out in the mailing list discussion that both "(Fictional)" and "(Imaginary)" were poor choices for long-term guidelines since "there is precedent for "fictional" fandoms to become real in their own right. (Castle & Jessica Fletcher's novels for example)."

Wranglers offered other suggestions for a disambiguating term to be used for either all Fiction-within-Fiction fandoms or those that lacked an easy solution (e.g., the attaching the author name to the Fiction-within-Fiction version of *The Princess Bride*). These suggestions included "in-universe," "parafandom," "metafandom," and "metafictional." Despite reaching a consensus that "metafictional" was the most accurate, specific term to describe these types of fandoms, wranglers still objected to its use given the likelihood of confusing users. "Meta" already had two common uses within TFR's community: metatags, such as "Marvel Cinematic Universe," under which users and wranglers collected multiple related fandom terms (e.g., "Iron Man (Movies)"), and meta,

174

which fans generally used to refer to a discussion of fandom itself or stories in which characters talked about their worlds as fandom. Wranglers had to consider authenticity in the classification system from the perspective of designers as well as typical users who would use these tags without the context of the wranglers' discussion. While wranglers agreed metafictional was authentic to the type of fandoms in this event, using it as disambiguation carried the risk of confusing users who would encounter it in search.

The second conflict between authenticity and effectiveness arose from the hierarchical structure of the curated folksonomy and the typical sense of hierarchy in classification systems. Though the curated folksonomy was not required to satisfy strict logical tests of class relationships (such as whether each subtag-metatag relationship could be written as "X is a Y"), wranglers tried to ensure that the hierarchical structure of meta- and subtags followed common sense. For example, a subtag to the metatag "Marvel Cinematic Universe" had to be part of that universe, and the subtags of that subtag had to pass the test of belonging to the "Marvel Cinematic Universe." Wranglers often examined larger hierarchical structures within the classification system for instances of "concept drift," a telephone-game effect whereby successive connections between tags stretched the "is-a" relationship between top- and bottom-level tags. The concern for the sanctity of the hierarchical relationship was not just conceptual; it was also functional. Searching or filtering by metatags, by default, returned works tagged with any of its subtags. Therefore, the thinner the connection between a metatag and all of its subtags, the greater the likelihood of irrelevant works listed in the search results for the metatag, negatively affecting the precision of filtering.

In the course of the discussion on Toni's question regarding subtagging Fiction-within-Fiction works to their source fandoms, some wranglers supported an alternative option, making both the source fandom and the Fiction-within-Fiction fandom subtags of

175

a new, common metatag. Some metatags of this type already existed on the archive. For example, "Sherlock Holmes & Related Fandoms" was an umbrella fandom over the Arthur Conan Doyle books, the Guy Ritchie films, and *Elementary,* an ongoing television series, among other works. Wranglers who supported parallel or "sister" relationships rather than meta- and subtag relationships between the source and Fiction-within-Fiction fandom argued for this solution based on precision in filtering: by making the two fandoms parallel, it would be possible for users to selectively browse and search across both fandoms or exclusively view one fandom at a time. The connection through the "& Related Fandoms" metatag would make searching across the conventional and Fiction-within-Fiction fandom tags easy and would signify to users the association between the two fandoms.

The wranglers who instead supported the original suggestion of subtagging the Fiction-within-Fiction fandom to the source fandom also argued that it would support filtering, though this solution would prevent users from viewing only those works from the source fandom: the Fiction-within-Fiction fandom would always be folded into search results. The advantage to filtering here was only in comparison to the two fandoms having no hierarchical relationship at all. Instead, the primary advantage of subtagging the Fiction-within-Fiction fandom to the source material fandom was in the authenticity of the classification relationship to the conceptual relationship between the fandoms. Fiction-within-Fiction fandoms had a part-whole relationship to their source fandoms: they were not parallel to their source fandom—they were within it. For wranglers and users who considered the hierarchical relationships in the curated folksonomy as representative of relationships between the represented concepts, this solution was the appropriate one. However, this solution would not allow the same relationship to play out in the user's experience of searching and browsing: though the Fiction-within-Fiction

fandom was dependent on its source for its existence, as a subtag users could isolate it in search, while users could not isolate the source fandom from the Fiction-in-Fiction fandom. Precision in searching favored only the smaller, dependent fandom.

Once the Fiction-within-Fiction event was live, with no wrangler consensus and therefore no fixed rule in place, wranglers overwhelmingly made the choice of hierarchy based on authenticity for the fandoms within their domain of responsibility. Some Fiction-within-Fiction fandoms, such as "Satan's Alley" or "Celestial Emporium of Benevolent Knowledge," had no relationship to their source fandoms. Wranglers made other Fiction-within-Fiction fandoms, such as "Wizardy Herbert," subtags to their source fandoms. I could find no cases of fandoms made parallel in the hierarchy, even when the existing tag structure would lend itself to this structure, as in the case of many books arranged under a single author.

SUMMARY

The concept of a work can be complex in any system (Bennett, Lavoie, & O'Neill, 2003; O'Neill, 2002; Peponakis, 2012; Renear & Dubin, 2003; Tillett, 2005), and the above examples reflect this complexity. In the context of fanworks, the community's focus on the source material works from which fans derived new creative works increased this complexity. In any system, works are a vital element of the organizational schema, especially as we conflate the notion of "work" with the "item" that circulates through the library. In TFR, works, as fandoms, took on an additional level of importance as they were the dominant attribute by which the system organized fanworks and wrangler tasks.

The above examples illustrate cases in which deciding the name or the boundary of a fandom touched on deeply held beliefs. In some cases, the wranglers as members of

the community held these beliefs themselves. In others, the wranglers stood in to champion beliefs they thought should be represented in the system. In some cases, the nature of the complexity extended beyond TFR itself, while in others the complexity was unique to the mission of TFR to collect and promote diverse fanworks. Wranglers worked through issues that did not have clear answers and that had histories of disagreement and division within the broader fanwork community and in society in general.

In the decisions above, wranglers created new solutions for complex wrangling tasks not addressed by TFR guidelines. In this and the previous findings chapter, the role of guidelines has been a recurrent theme. In the next and final findings chapter, I explore in greater detail how wranglers understood guidelines, including how they handled decisions they believed were exceptions and how they developed guidelines to handle new cases. This change in focus will illustrate the role that guidelines played in design decisions and how wranglers diverged from the view that classification is an exercise in rule-following.

What's in a Guideline?

In this chapter, I describe how wranglers interacted with guidelines and, in particular, how wranglers understood guidelines' role in design decisions. The extent to which guidelines determined wranglers' design decisions is particularly important to distinguish between classification as rule following and classification as a personal, creative act. As the examples in this chapter illustrate, guidelines had an important but not determinative role in wranglers' design decisions. While wranglers valued the guidelines and were thoughtful in developing and revising guidelines, they primarily followed their domain knowledge with regards to their fandoms.

As with the previous thematic chapters, this chapter on guidelines follows the experiences of wranglers when their interactions with guidelines were complex. The first section, "Guidelines at TFR," reviews how wranglers encountered and defined guidelines. The second section, "Violating the Guidelines," examines cases in which wranglers felt that the guidelines did not support their tasks and how they justified acting against the guidelines. All participants reported some trouble with following the guidelines, though they differed in whether they interpreted divergence as a problem or as an expected limitation of the guidelines' generalizability and reach. New wranglers were more likely to treat the guidelines as strict standards. Wrangling experience presented more and more instances in which a wrangler decided to violate or ignore a guideline. The third and final section, "Making the Guidelines," examines how wranglers revised guidelines with a focus on how wranglers described good rules. Alongside the performance of their daily design work, wranglers eventually participated in the development and revision of guidelines, reflecting on their experience to provide input.

This final section describes how wranglers contributed to guidelines and what their discussions reveal about the concept of guidelines.

GUIDELINES AT TFR

While official documentation used the term "guidelines" for all guidance, from the wording of general principles to the standardized forms of specific tag types, wranglers sometimes used the word "rules" to refer to the same. Figure 16 illustrates the relationship I observed as being the dominant understanding among wranglers. Principles informed guidelines, which rules instantiated in shorthand. Rules were the most concrete form of guidance while principles were the most abstract.
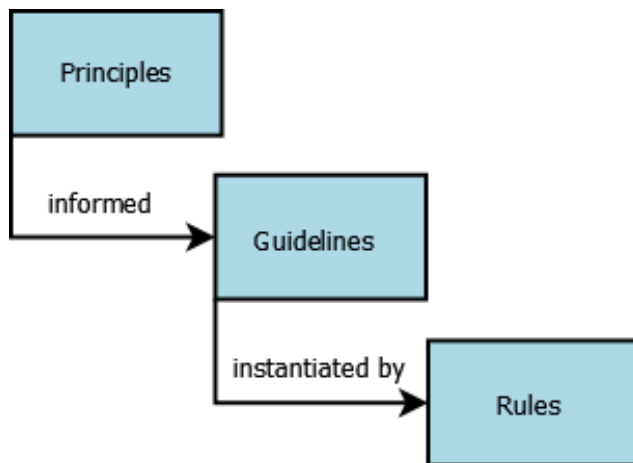
Figure 16 Relationship between principles, guidelines, and rules

Principles had the greatest authority and were the basis from which wranglers developed guidelines. For example, wranglers commonly cited the "Rule-of-3," which referred to the minimum threshold of concept use before wranglers intervened to create a canonical tag. The Rule-of-3 did not apply to every situation but it was a common heuristic for prioritizing actions according to the principle of "user primacy," which stated that

wranglers' decisions followed rather than dictated tag usage on TFR. In other words, the Rule-of-3 was a simple shorthand that prioritized wrangler attention toward tags that met a minimum standard of usage and would offer wranglers examples of user tagging choices upon which to base their design decisions.

As an example of how principles, guidelines, and rules figured into a single decision, I describe how and when wranglers created new freeform tags for episodes. The principle of "consistency," which instructed wranglers to standardize tags across fandoms, necessitated the guideline that mandated formats for episode tags (ie., Episode: s##e## EPISODENAME), while the Rule-of-3 suggested that wranglers not create a canonical episode tag until three users had tagged for it on three separate works. Principles reflected the philosophy of the curated folksonomy, and included very general concepts such as "consistency" and basic design practices such as "metatagging," the building of tag hierarchies. Guidelines provided standard forms for each tag type and provided examples for common phenomena, such as the format for episode tags, above. Though wranglers frequently used the term "rules" to refer to the guidelines, more commonly rules were simple shorthand for implementing the guidelines, such as the Rule-of-3.

Wranglers found guidelines in two sources: the tag wrangling guidelines available to the public and training materials within the password-protected wiki. Wranglers familiarized themselves with the public-facing guideline documents as part of the training process. These documents covered each tag category (Fandom, Character, Relationship, and Freeform) using several prototypical examples covering common cases. The training documents and peer wranglers explained these guidelines to new wranglers in more detail. New wranglers also performed related tests to assess their ability to implement the guidelines. The public-facing and internal, training documents predominately dealt with

guidelines, though the public-facing documents also gave short definitions of principles. These documents identified most, but not all, of the principles wranglers identified as informing guidelines. For example, wranglers spoke of "wrangler discretion" as if it were a principle though the guideline documents did not identify it as such.

In most cases, wranglers' interactions with the guidelines were helpful and uncomplicated. Through the diary protocol, I prompted participants to record instances of "straightforward" and "complex" wrangling tasks, leaving it to the participants to define these categories through examples. Many straightforward tasks involved checking the guidelines for an infrequently encountered tag type and finding an easy-to-implement answer, such as in these diary excerpts:

> forgot what to do with named non-fandom specific original characters, referenced wrangling guidelines, dealt with ~10 of those (Victoria, Diary)

> The last one is a smushed pairing name – I don't get many of those, so I have to look up what to do with it in the [TFR] FAQ [Public wrangling guidelines]. (Marina, Diary)

The predominant attitude of the wranglers regarding the guidelines was appreciation; guidelines supported wranglers' work by providing template term and term relationships for established but irregularly encountered tag phenomena. In rarer cases, the guidelines were part of "complex" tasks, as wranglers perceived that the solutions and requirements the guidelines presented did not fit their current problem. These rarer cases are the subject of this chapter because examining them reveals how wranglers understood rather than applied the guidelines. It was through these complex tasks that wranglers reflected on the purposes of guidelines and the factors that conflicted with the application of guidelines. In the next section, I use examples of complex decisions in which wranglers chose to violate the guidelines to illustrate the wranglers' practice of applying their own judgment rather than interpreting classification design as rule following.

Alba, a novice wrangler—one who joined two months before she began the diary—described the experience of interpreting a revised guideline for her first fandom. Before describing how Alba understood the guideline in terms of her responsibilities to her fandom, I explain the revised guideline itself, which instructed wranglers to remove meta/subtag connections between character tags denoting groups (e.g., The Rolling Stones) and character tags denoting members of that group (e.g., Mick Jagger). The text of the new guideline, finalized two weeks before Alba's diary, explained:

> This is because group canonicals are used to indicate the entire group. Subtagging individual characters would mean that users looking for works featuring the entire group would also see works featuring only one or two characters in their results. For similar reasons, groups are also not subtagged to the canonical characters who make up the group. (Internal guidelines)

The revised guideline reflected the view of the wrangling team that precision in user retrieval was negatively affected by these tag hierarchies. In other words, hierarchical relationships between group and group member character tags created search results that included irrelevant works. As with all guidelines, this guideline was intended to apply to all fandoms' group and group member character tags.

Alba did not believe this rule worked for her fandom. Figures 17 and 18 use the example of The Rolling Stones to compare the hierarchy of tags Alba encountered (Figure 17) with the form they would become under the new guideline (Figure 18).
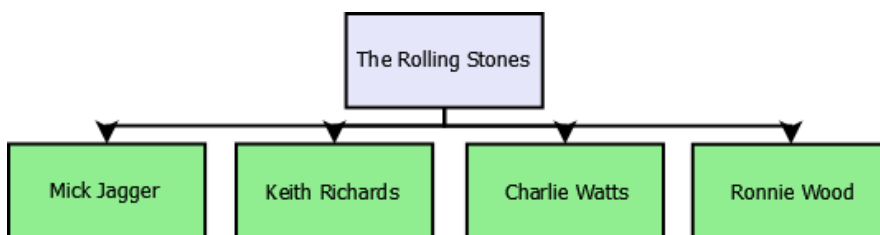
Figure 17 Group-Character Tag Hierarchy
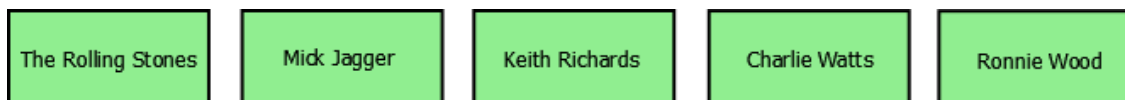
Figure 18 Group and Character Tags without Hierarchy

In the original structure, which Alba wished to keep, clicking on "The Rolling Stones" would return all fanworks tagged with that string (and synonyms such as "Rolling Stones") as well as all fanworks users had tagged with any band member. Without the hierarchy, as in Figure 18, "The Rolling Stones" would only return fanworks tagged with that string and its synonyms. The tags would no longer support a filtered search that returned all fanworks that contained any member of the band.

Alba's experience as a TFR user who read and wrote in her fandom led her to believe that the links between the group and character tags were consistent with how users in the fandom tagged their works and were important to retrieval:

> Alba: We had the discussion about group tags—meta and subtags—earlier this month, I think, and in that discussion it was decided that subtags should be de-linked from metatags where necessary. I wasn't sure which group tags should be left alone and which one should be de-linked. […] I feel like, at least in [this fandom], these tags need to have the characters attached otherwise there's no point.
>
> Interviewer: What is their [the other wranglers'] point of view, from your understanding of it?
>
> Alba: Like, people are using those tags to refer to specific characters, but then only really referring to one or two of that group. So it would be very

inconvenient/annoying to not find the fics [stories] you wanted from that tag. So I get it from that perspective, but I still can't really get on board with it. (Alba, Interview)

Alba's fandom was a large fandom she shared with more experienced wranglers who agreed with the guideline change. Knowing that her cowranglers had a different point of view, Alba chose not to implement the new guideline and left the tags alone for her cowranglers to make the change or convince her of their point of view. In this instance, the wrangler's choice was to make no change at all, deliberately in violation of a guideline that mandated changes.

Wranglers with more experience than Alba encountered similar problems and described the discrepancy between specific guidelines and their fandoms in more positive terms, without the worry and conflict Alba experienced. Where Alba deferred making changes until she could understand the guideline better, other wranglers declared that they could violate guidelines that did not fit the tagging practices of their fandoms. Catherine, a wrangler with three years' experience, described a role-playing game fandom where the common experience of source material necessitated a divergence from the character tag guidelines[6]. Specifically, Catherine referred to the guidelines for Original Characters, a special type of tag designating characters which were the invention of the fanwork author, rather than derived from the source material.

> I actually, I made a decision a long time ago, about how people want to find fic in this fandom. And so I know that it actually goes against guidelines but I do it anyway. I created an Original [Fandom] Character tag, and I mostly just syn them to that, because people who are going to be reading in this, people who are writing in this fandom, are not going to have trouble finding their own fic again, right? People who are wanting to read stuff, in terms of being able to find stuff, I think that it was a more useful way of organizing in a fandom that is mostly Original Characters. Instead of just send them all away. […]

---

[6] In role playing games such as *Dungeons & Dragons*, players create characters within the game's setting. While there are pre-written characters encountered in many players' games, the bulk of storytelling occurs through the player-created, rather than source material, characters.

It gives people something to use too if they don't want to create, if they don't want to use "Original Characters" or if they don't want to necessarily tag for individual people. I don't think it makes sense for every fandom. But for fandoms where you do get people where the bulk of people's interactions with it as a canon outside of fic is through Original Characters, I think it makes sense. (Catherine, Interview)

Like Alba, Catherine believed that the relevant guideline did not work for her fandom. Whereas Alba chose not to implement the revised guideline, Catherine created a fandom-specific rule that conflicted with existing guideline. Instead of sending a character such as "Araggul Hardheart (Original Character)" to the Original Work fandom, as per the guidelines, Catherine linked it to an umbrella term within her own fandom. Figure 19 illustrates the organization of tags according to the guidelines (left, within the fandom "Original Work") and the organization chosen by Catherine for her fandom (right, within Catherine's fandom).
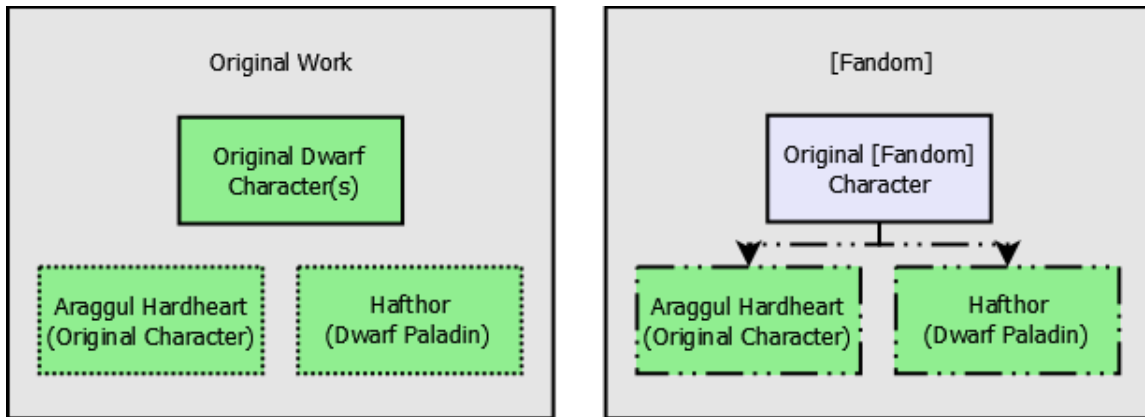
Figure 19 Location of fandom-specific Original Characters

According to the guidelines, as on the left, wranglers designated Original Character tags as belonging to the Original Work fandom, but would leave them in an Unfilterable state—unconnected to each other and retrievable only by their exact string. Wranglers did not make Original Character tags canonical, nor did they make them synonymous to a generic canonical character tag such as "Original Dwarf Character(s)." For example, a

user clicking on "Araggul Hardheart (Original Character)" would find only works tagged with that exact string. In Catherine's process, she kept Original Characters that occurred in her fandom and that were sufficiently specific to her fandom within that fandom and grouped them as synonym tags of the canonical tag "Original [Fandom] Character." In this case, and as illustrated in the right half of Figure 19, clicking on "Araggul Hardheart (Original Character)" would return works tagged with that exact string as well as works tagged with "Hafthor (Dwarf paladin)" and every other tag within that set, displayed as search results for "Original [Fandom] Character." Making the users' character tags synonyms of the generic character tag made it more difficult for users to find a set of works about a given named Original Character, as they could not be distinguished from others in tag filtering, but made it easier for users to find all works featuring Original Characters from this fandom. Catherine made a measured trade-off between these two filtering scenarios and chose to diverge from guidelines to create a structure that would serve the retrieval needs of users in her fandom.

The above two examples show that, when wranglers diverged from guidelines, it was because they feel that the guidelines did not serve their specific fandoms. In many cases, wranglers described this discrepancy between TFR-wide guidelines and the needs of their fandoms as fandom "weirdness." In the rest of this section, I provide further examples that illustrate the role fandom weirdness played in wranglers' justifications for guideline divergence.

Often, wranglers' declarations they would not follow a particular guideline followed my interview prompt for wranglers to tell me about their set of fandom assignments. Catherine offered that she "wrangled a lot of weird things" and we spent most of our time—as was the case across participants—discussing the "weirdness" of specific fandoms. Brigid and her cowranglers, similarly, observed atypical tagging

practices within their fandom and decided on a fandom-specific wrangling rule that diverged from the guidelines:

> The biggest argument with those group characters is: according to the guidelines you should just canonize them as either characters or freeforms and make the other one unwrangleable because they will look exactly the same. But we did have a long conversation about how they are actually different usage. (Brigid, interview)

The guideline to which Brigid referred stated: "Do not canonize tags that are duplicating another format of tag (a character name, a pairing name, a fandom name)" (Internal guidelines). Brigid and her cowranglers argued that, for example, "Marine(s)" as a character tag had a sufficiently different meaning than "Marines" as a freeform tag; the former identified one or more included characters, while the latter referred to their organization. This phenomenon was so common in Brigid's fandom as to necessitate creating a fandom-specific wrangling rule within their internal documentation on the organization's wiki (along with clarification on how to handle irregular plural forms):

> We have a bunch of character tags that refer to groups of people. There is also a lot of overlap between group Character tags and group Freeform tags. While it is proper to have both, they should be formatted consistently. The rule of thumb is that group Character tags exist to refer to the people, while the group Freeform tags exist to refer to the general concept of that group. So most group tags will follow the following format:
>
> - Character: [Singular group name](s)
>
> - Freeform: [Plural group name]
>
> (Brigid's Fandom wrangling wiki documentation)

Brigid and her cowranglers acknowledged that the guideline disallowing duplicate tags across categories was correct; they justified their divergence from practices in other fandoms by stating that the terms did have sufficiently different meanings to their users. As with Catherine's choice to create a fandom-specific rule for Original Characters that

contradicted the guidelines, Brigid and her cowranglers created a fandom-specific rule that made an exception for the guideline on duplicate tags.

If Brigid and her cowranglers had followed the guideline, they would have had to choose which tag category version of the group name to designate as the canonical: either the group name as a character tag or the group name as a freeform tag. The version of the tag in the other tag category would have remained fragmented. For example, if they chose "Marines" as a character canonical and not provided a freeform canonical, no variations of user tags of "Marines" in the freeform field would collocate in filtering nor would they appear in autocomplete. Stories featuring Marines as characters would be more visible in filtering than stories about the Marines as an organization. Similar to what Catherine did alone, Brigid and her cowranglers weighed the effects of each approach on filtering and made a choice that diverged from guidelines.

Another experienced wrangler, Tamara, discussed how her familiarity and passion for the source material for her fandom motivated her to be very selective in designating some popular concepts as canonical tags. In her diary, she documented various user-created tags of the type "Student Activist [Character Name]." The source material for this fandom featured young characters engaged in student activism at a troubled time in their country's history. In her diary she annotated the student activist tags with the comment "THAT'S CANON, so forget it" (Tamara, Diary). During the interview, she explained that the relationship between source material canon and TFR tags necessitated or justified a divergence from wrangling practice. Specifically, Tamara stated that she would not follow the practice whereby wranglers made canonical tags after multiple users tag for the same concept:

> This is something that is very specific to the fandom. And it really annoys me
> when people do it and I'm the wrangler, so, it's sort of within my discretion to

discourage people from doing [it] through my tagging policies, as long as I'm not breaking the guidelines. […] And so, if people are tagging for something that is definitely canon in the character tags but they're tagging it like it's a, this is a specific version of the character, I can use my wrangler discretion and work out my frustration and don't tag it. And [Character Name] is canonically a student activist who spends all his time in coffee shops so I am not making a special tag for him to be a student activist. […] And so I don't want to encourage people to look at the drop-down [autocomplete] and say oh the canon versions are the special exception. As a fan, I want to encourage people to think of the canons versions as the default, which it should be. (Tamara, Interview)

Here, Tamara named a principle only implied in the accounts of Brigid and Catherine: wrangler discretion. Tamara interpreted her divergence from the Rule-of-3 and the principle of user primacy as wrangler discretion within guidelines. Even though the variant uses of "Student Activist [Character Name]" met the minimum requirements for creating a canonical tag, Tamara left the tags unfilterable (i.e., without a canonical) out of respect for the source material.

The guidelines gave Tamara the room to make this interpretation by stating that the Rule-of-3 designated the minimum rather than sufficient number of tag uses before a wrangler intervened to designate a canonical tag. Still, Tamara described her choice as a rebellious one and one that enacted a principled and even contentious attitude within the community. TFR's principle of "user primacy" was clearly at odds with Tamara's decision, as it dictated that wranglers not pass judgment on user tags. By selectively implementing the guidance of the Rule-of-3, Tamara avoided reinforcing what she saw as incorrect tagging behavior in her fandom. For user interactions with the system, Tamara's choice meant that the variant tags for this concept would be fragmented and would not be collocated in search, nor would any version of the tag appear in autocomplete in user search or for posting new works. If someone wanted to find a story in which this character was a student activist, their only choice would be to filter by his character tag,

returning all the works where he appears as a character—in which, according to Tamara, he would often have the characteristics of a student activist.

Though Tamara's example was an extreme one, her explanation that she violated guidelines to best serve her fandom was consistent with the views of her peers. Another wrangler, Amelia, explaining why she sometimes made different decisions for similar tags in different fandoms, summarized the point of view common to experienced wranglers:

> At the end of the day, while we do strive to standardize with [TFR], our first duty is to our fandoms. Which is why there's a proviso on the guidelines that [TFR] has some set ways, but at the end the final decision is with the wrangler of a given fandom. (Amelia, Interview)

Amelia here described a dichotomy that was at play each time wranglers chose to violate the guidelines: that between a wrangler's duty to TFR and to her fandoms, the latter was most important. Consistently, wranglers made the same interpretation as Amelia; while they shared common guidelines, these guidelines were in service of, not meant to be counter to, the needs of their fandoms. When wranglers presented divergent practices as fandom-specific, they did so not by criticizing the guidelines as flawed but as inappropriate to their given fandoms. As wranglers gained experience of more fandoms—directly, through their own assignments, or indirectly, by seeing others' dilemmas on the mailing list or in the chatroom—they became more confident in setting aside TFR-wide guidelines in favor of fandom-specific rules.

In addition to incidents in which wranglers chose not to implement a guideline, to implement a guideline selectively, or to create a contradictory fandom-specific rule, it is important to note what I did not find in my data: incidents in which wranglers believed a guideline to be inconsistent with their fandom and its tagging practices but implemented it anyway. Guidelines could be helpful, as in the typical cases in which wranglers

checked the guidelines for a reminder of a less familiar case, or they could be violated, as in the cases above, but guidelines did not force wranglers to make a decision they judged to be at odds with the needs of their fandom. Instead, wranglers would take one of the paths illustrated above or they would reach out to other wranglers, through the mailing list, the chatroom, or the staff email account, for help in interpreting the guidelines in a productive way—or for encouragement to ignore the guidelines in a given case.

Though the wranglers occasionally referred to the guidelines as "rules," and described them as stating what they as wranglers "should" have done, the team's choice to call these "guidelines" was a meaningful one. The guidelines here were similar to standards or rules in other classification settings, such as the NISO standard for controlled vocabularies (National Information Standards Organization, 2005) and the Anglo-American Cataloging Rules (Joint Steering Committee for Revision of AACR et al., 2003). Standards such as NISO have included the statement that they are "recommendations," leaving open the possibility for designers to make other choices (National Information Standards Organization, 2005, p.1). Hoffman notes that such divergences are rare in settings where designers perceive pressures to conform to standards, such as in cataloging work where standards facilitate speedy work and interoperable catalog records (Hoffman, 2009).

At TFR, there were some consequences for violating the guidelines. Staff members occasionally performed reviews of fandoms, which they termed "wrangler check-ins," during which they would provide private feedback, including whether they observed tags that were not consistent with guidelines. This process was a dialogue between the staff wrangler and fandom wrangler, and gave the latter an opportunity to explain divergences and negotiate compromises, if necessary. Similarly, complaints from users about "incorrectly" wrangled tags were routed through volunteers on the Support

team; the volunteers communicated with individual wranglers to determine whether there was a mistake to be fixed or whether a reported inconsistency was a wrangler's intentional and principled decision. In the latter case, the Support volunteer would then relay the wrangler's rationale to the user.

In more extreme cases, repeated divergences from a guideline or requests for staff to clarify guidelines initiated a guideline review process. The following section details how wrangler experience and the growth of TFR would then contribute to the development and revision of guidelines.

MAKING THE GUIDELINES

In this section, I show how wranglers created and revised guidelines. Despite the accounts in the above section, it is important to note that wranglers typically acted consistently with guidelines and used them as resources to simplify their design decisions. Therefore, a wrangler's contribution to a guideline was also a contribution to future design decisions. I first describe how the guideline creation and revision process changed during my time at TFR and then examine in detail the contributions wranglers made to this process.

Ad Hoc Guideline Development

The early development of wrangling practice involved the simultaneous development of standardization in tag formats and of the technical features wranglers requested from the technical development team. As an illustrative comparison, while wranglers were settling on an alphabetical-by-family-name order for relationship tags (e.g., James T. Kirk/Hikaru Sulu, here "K" comes before "S"), they were also encountering the technical limitation of the initial 42-character length of fields, so that users and wranglers could not enter longer relationship tags (James T. Kirk/Leonard

"Bones" McCoy/Montgomery "Scotty" Scott/Spock/Hikaru Sulu) into the system. When wranglers reported that this technical limitation was not compatible with the needs of the curated folksonomy, the technical development team made code changes allowing for a 100-character length for tag fields. Across the history of the wrangling project, the development and revision of guidelines was embedded in the growth of the technical infrastructure, including changes that wranglers requested and changes that wranglers agreed to in order to minimize the burden that wrangling actions put on the repository's server. The technical change to character length for tag fields is an example of the former; a guideline change to regulate when wranglers would make changes to very large (>5000 works) tags is an example of the latter.

The process by which wranglers created and revised guidelines had itself changed, from an ad hoc process to a more systematic one. A wrangler involved in the first days of the curated folksonomy project and its guidelines stated, "slowly we started setting up standard tag formats for things" (Tifa, Interview) with a priority of consistency across the repository. As the collection grew and more wranglers joined the team, the norm became for changes to be initiated by wrangler questions. Wranglers would approach staff with a question on how to handle a case from their fandoms, as Tifa explained:

> Guidelines develop as 'issues' come up - something doesn't fit quite right with the existing guidelines, so then people ask, "What about X?" and the committee has to determine if it's something that really NEEDS a guideline added or if it's something the wranglers can determine on their own. (Tifa, Interview)

The latter cases, in which the conflict was resolved an as exception within the fandom, were consistent with the examples illustrated earlier in this chapter. It was typical in wrangling work for wranglers to find "weirdness" in a fandom—based on peculiarities of the source material or fan activity around it—that justified a divergence from the TFR-

wide guidelines. In some of these divergences, wranglers acted autonomously, as in Tamara's appeal to wrangler discretion. Other wranglers, by Tifa's account, developed these divergences and adaptations after consultation with and with the formal permission of staff wranglers. If the committee—the staff wranglers—agreed that the wrangler's question needed to be addressed in TFR-wide guidelines, revision occurred through discussion in staff meetings, open to the entire wrangling team through the chatroom and its logs. As the curated folksonomy grew, as wranglers observed the same problems more often, and as these problems became interconnected and defied individual solutions, the wranglers replaced this ad hoc process with a more systematic process.

Systematic Guideline Revision at TFR

Beginning in July 2015, the typical system for revising guidelines ad hoc and as needed transitioned to a systematic review of persistently problematic guidelines. The staff members who organized these revisions brought several guidelines into a formal discussion process for all wranglers. Staff members facilitated these discussions in subsequent phases marked by separate threads: Questions/Brainstorming, Compromise/Drafting, and Conclusion/Implementation. The guidelines staff presented to the wrangling team as a whole affected multiple fandom domains, would involve substantial rewriting of guidelines, and could require the re-wrangling of many tags in the curated folksonomy. For example, guidelines that impacted the hierarchical or "tree" structure of the curated folksonomy were particularly urgent objects for revision. In this section, I do not provide names for quotes from the mailing list, but only designate their role—a staff member or a wrangler. The former were wranglers who had volunteered for and been given extra responsibilities including facilitating guideline revision discussions.

To show how the conceptual, technical, and retrieval impacts of these structures are interconnected, I return to the guideline from Alba's wrangling experience, prohibiting metatagging group character tags to group member character tags. Previously, I described how the original structure would return all works tagged "Mick Jagger" when "The Rolling Stones" is searched, illustrated here as Figure 20.
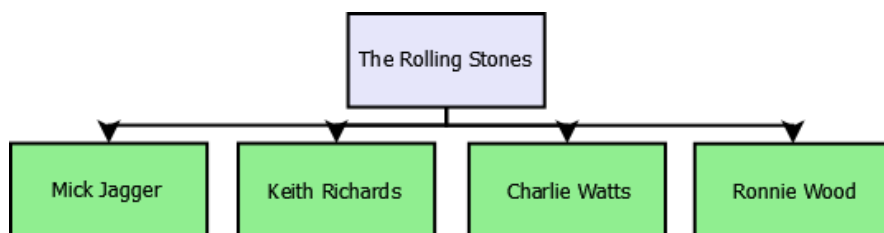
Figure 20 Hierarchical tag structure

From the technical, database-level interaction of TFR, this result occurred because the server would index every instance of "Mick Jagger" with "The Rolling Stones" as well. Therefore, each metatag relationship created redundant indexing changes within the database, adding burden to the server while adding dubious filtering capabilities for searchers. Though, conceptually, the hierarchy was accurate, the filtering functionality it and other trees triggered made such a burden on server resources that they slowed the site for all users. The impetuses for making systematic guideline changes were therefore an aggregation of conceptual problems with the guidelines and pressure to minimize the negative effect of the curated folksonomy on the speed of server requests.

In the following subsections, I describe how the guideline revision discussions revealed wranglers' attitudes towards guidelines, particularly their shared idea of what made for a "good" guideline, what they believed the limitations of guidelines to be, and

what they indicated were valid types of evidence and authority in contributions to guidelines.

Good Guidelines

In the daily experience of wranglers, a good guideline was one that was clear and helpful but not overly ambitious; good guidelines did not present themselves as being applicable to more cases than those for which wranglers developed them. For example, a frequently cited guideline in participant diaries described how to process original or generic character tags. The guideline was clear in that it included five named variations of the phenomenon, all with at least two illustrative examples, stating "tags like [x] or [y] go to…". Wranglers could check the tag in question against the list and resolve the tag that spurred them to consult the guideline. The guideline was helpful in that it offered concrete solutions and it taught wranglers the terminology or jargon applicable to the phenomenon at hand. In this case, wranglers learned that they should refer to character tags such as "a stray dog" as "Generic Character Tags." Good guidelines provided clear processes or examples to follow, gave wranglers helpful tools—such as jargon—for more complex cases, and established suitable limits to their application. For example, in the case of the original or generic character tag guideline, the guideline noted that special cases for recurring "fanon" (fan-canon) Original Characters should be handled differently than for other Original Characters.

The discussion around systematic guideline revisions made more explicit the wranglers' ideas of "good" guidelines through the parameters staff wranglers set for the discussion and through the discussion itself. I use examples from staff's instructions to illustrate how these experienced wranglers defined a good guideline. For example, instructions from the staff managing the process identified some elements of a good

guideline. A staff member introduced the second phase of one guideline revision discussion, in which wranglers synthesized the open discussion feedback into a solution, as follows:

> During this period, we will ask everyone who participates to come to a rough consensus, if possible. That doesn't mean everyone will agree perfectly or have exactly what they want as the final result. Instead, it means that everyone will be able to work with the result and continue to wrangle without major impediment. Our goal is a net improvement in the situation for all wranglers and users. (Staff member, Mailing list discussion)

From the staff's perspective, good guidelines were not necessarily those that were correct but those that were good enough to support the ongoing work of the team. This idea of incrementally—if imperfectly—improving the curated folksonomy was made clearer by the background or problem statement for each guideline revision. In introducing one guideline revision relevant to the hierarchy of "Alternate Universe" (AU) tags, the staff member's statement included the following constraint:

> The AU Tree is enormous and complex. Please keep your discussions restricted to this particular aspect of the AU tree. Other aspects will be discussed in upcoming [guideline revision discussions]. (Staff member, Mailing list discussion)

This constraint situated the discussion within a longer process of revision and improvement. This perspective was consistent with experienced wranglers' impression of the large scope of these problems and helped to introduce this perspective in newer members.

The idea of incremental improvement also arose as the community pursued ethical and contemporary terminology through guideline revisions. New or revised guidelines to make the curated folksonomy more consistent with progressive language were not final but continuous and in parallel with social and linguistic change. In the case of a proposed guideline revision to provide a replacement term for "American" when

meaning "from/pertaining to the USA," the problem statement noted that the current practice:

> …shows a lack of awareness of the debate surrounding both the term and the feeling of belonging and ownership that several L[atin]A[merican] countries have towards America, the continent. It doesn't reflect well the org's values or its commitment to diversity and inclusivity. (Staff member quoting internal feedback, parentheticals in original, Mailing list discussion)

The discussion around this guideline revision was particularly challenging, as it involved resolving an ambiguity or injustice not perceived by all wranglers. The communities affected by this practice, namely United States Americans and Latin Americans, were members of the TFR community, both as users and wranglers. This guideline discussion was presented as an opportunity to ameliorate—through a very specific but representative change—a United States-centric bias in the system.

These excerpts represent the framing that staff members, in their role in organizing and facilitating guideline revision discussions, provided around guidelines. The following two subsections more directly address the framings of guidelines that emerged through the discussion and from the contributions of the wrangling team. First, I discuss how wranglers understood guidelines as having limitations. Second, I discuss what wranglers perceived to be legitimate evidence and authority in decisions around developing and revising guidelines.

Limitations of Guidelines

A prominent theme in the revision discussions was that there were more and less productive ways of thinking about guidelines. Experienced wranglers, as in the first half of this chapter, perceived guidelines to be limited and negotiable; having encountered a guideline that contradicted their judgment, they would assert that their case was an exception. Guidelines that wranglers frequently found helpful, as in the above section,

made claims to their own limitations by identifying special cases to which the guideline did not apply. Team-wide discussions such as the guideline revision discussions contributed to this view of guidelines as limited and negotiable. Witnessing and participating in the messy, compromise-laden, and tentative process of establishing guidelines revealed to wranglers the situated nature of guidelines.

In guideline revision discussions, wranglers continually established the limitations of guidelines. One way in which wranglers established and highlighted guidelines as limited was by scoping the guideline revision discussions. Wranglers frequently negotiated the boundaries of the guideline under discussion: which problem cases revealed flaws in proposed solutions and which were beyond the scope of the current discussion. I explored an extended example of this type of negotiation in the Fiction-within-Fiction section of the "What is a Work?" chapter. Similarly, during the discussion over how and whether to introduce a guideline mandating how "American" is used in designating the United States of America, a wrangler for affected history and folklore fandoms requested that the discussion not seek to address these fandoms. She first explained the complicating factors for these fandoms, concluding:

> So as the person who wrangles [this subset of fandoms], I'm going to very nicely ask that they not be brought back up as part of this discussion. Suffice to say that they will eventually be changed, but that isn't going to happen right now. The decisions made now will however be taken into account when these fandoms names finally do get altered. (Wrangler, Mailing list discussion)

This discussion contribution served two functions: it streamlined the conversation by excluding particularly challenging cases from the scope of the conversation and it preserved the autonomy of this wrangler in managing her own fandoms. However, other wranglers pushed back at this particular attempt at boundary-setting:

> In order to make things easier for more wranglers […] I do think it's important to try and keep things consistent and limit the number of exceptions we have to help

make it easier for wranglers to learn our (complicated!) guidelines and be able to follow them more consistently. (Wrangler, Mailing list discussion)

The latter wrangler expressed concern that continually keeping the one set of problematic fandoms beyond the scope of guidelines made the guidelines too difficult to follow. While she recognized that some fandoms have more complicating factors, the effects of waiting to address them were too great to ignore. The stated goal of the proposed guideline—to decenter United States of America-centric language—was particularly relevant to the fandoms the first wrangler was trying to exclude. The second wrangler's point received support and ultimately decided the basis for the scoping of the new guideline.

Another way in which wranglers established the limitations of guidelines was to invoke wrangler discretion and case-by-case analysis in revision discussions. The prevalence of wrangler discretion and case-by-case judgments in discussions of TFR-wide guidelines limited the authority of guidelines. Some wranglers used the notion of case-by-case judgments to refute the creation of a new guideline, claiming that too many instances would become exceptions:

I don't think we necessarily can or should ban the use of the term [American], as it is the accepted term for a variety of things (American Civil War, for example), and where there is no good alternative (which I seem to recall some historical examples of from the previous discussion) and it seems like a case-by-case kind of situation. (Wrangler, parentheticals in original, Mailing list discussion)

In the above wrangler's suggestion, the proposed guideline for how to create canonical fandom tags would instead become a suggestion or warning: If possible, avoid the use of ambiguous or contested terms for regions, such as "American." In other cases, wranglers advocated for a guideline that suggested a particular change while acknowledging that wrangler discretion was primary. The following wrangler suggested that, from experience

in a similar fandom, she believed the guideline would work for one of the fandoms under discussion, and included the following statement:

> I think that this is the kind of thing that is best left up to the discretion of that fandom's wrangler(s), as they would know what makes the most sense for that specific fandom. (Wrangler, Mailing list discussion)

This wrangler agreed that it is appropriate to consider the mentioned fandom within the scope of this guideline while leaving open the possibility that its wrangler would ultimately decide to implement a different solution.

Evidence & Authority in Guidelines

The above example reveals something about general structure of guidelines: they were limited in scope and they were subordinate to a wrangler's judgment of individual cases. Through discussion, wranglers pressed back against an idea that guidelines were truly universal or above wrangler judgment. Wranglers also indicated judgments in what was appropriate input to a guideline. In this subsection, I use excerpts from the guideline revision discussions to illustrate what wranglers presented as valid inputs to guidelines, specifically the types of evidence and authority they brought to bear.

In presenting claims and requesting the feedback of others, wranglers indicated what sources of expertise they deemed valid. In later instances of guideline review discussions, staff made explicit one notion of valued expertise with the following statement:

> This discussion will address a wrangling issue that may not be familiar to you. If you're a new wrangler, please email staff directly if you have questions about what this change means, to avoid derailing the conversation into wrangling fundamentals. Please let wranglers who have encountered this situation take the lead on discussion. (Staff member, Mailing list discussion)

In including this instruction in the framing language of guideline revisions, the staff indicated that wranglers should give the expertise of hands-on familiarity with the phenomenon top priority in these discussions.

One way in which wranglers indicated authority and expertise was through the metaphorical switching of "hats." Throughout guideline revisions discussions (and indeed in many team-facing discussions) wranglers with multiple sources of authority or expertise would state which "hat" they were wearing at any time, such as by adding "staff hat off" to their signature line. This practice helped to indicate to others in the discussion the authority or expertise they intended to bring to bear in their claims. The normative practice that emerged was to separate authority over the group from expertise over the subject; when wranglers-who-were-staff contributed a message to manage the guideline revision discussion, such as by reminding participants of the timing of the phases of the process, they wore their staff hat. When wranglers-who-were-staff contributed to the discussion itself, such as by contributing problem cases or proposing solutions, they took off their staff hat to wear only their wrangler hat.

While organizational authority was explicitly separated from subject expertise, other types of authority or expertise were invoked—or worn—as valid. The most persuasive form of expertise was that of a wrangler's experience with their own fandoms. Wranglers used their experience with an affected fandom to describe the impact of a change or offer an insight from a parallel case. The following wrangler contributed the latter type, describing how a practice in film might apply to historical fandoms.

> Bringing to this discussion of the Revolutionary period my two cents as a film wrangler, we don't typically disambig [disambiguate] a property until it has a sequel. Until there's another one, I don't think the fandom "American Revolution" needs disambiguation. (Wrangler, Mailing list discussion)

This wrangler used the established wisdom from one section of the site—films—to introduce a possible rule for the more complex historical fandoms.

Fandom wrangler expertise was not only offered, as above; it was often requested. Wranglers participating in revision discussions often tempered their own suggestions with admissions that they did not have hands-on experience with the relevant fandoms and would encourage those with that expertise to contribute. In the following excerpt, a wrangler explained what kind of insight they expected a fandom wrangler could offer:

> My inclination in all these cases is to defer to the wranglers who wrangle fandoms in that genre. Those wranglers know what's most clear to their users and are familiar with the issues involved. So I'd primarily look for concerns from those wranglers. (Wrangler, Mailing list discussion)

Comments such as this one were especially common when the fandoms in question had complexities known to other wranglers by reputation. Wranglers continually made room for wranglers with immediate experience of a phenomenon to provide their insight before trying to present solutions.

The most commonly invoked form of evidence or expertise was the wrangler's shared role as a TFR user. This dual-nature of designer and user was particularly useful when it revealed the interaction among fandoms in TFR. In response to a proposal that "role playing game" be used to disambiguate table-top and text-based game genres from their video "RPG" counterparts, one wrangler offered the following objection:

> I'm going to have to disagree about RPG being used only for video games and Roleplaying Game only for tabletop games. While I wrangle mostly video games, I play both and in my experience, this simply isn't true. There's really no distinction in the terminology and these days, especially, RPG is often used for both - even by people developing the games in question. So I don't really understand how this change in disambiguation would work; it seems to me that it would be extremely confusing. I sure wouldn't understand the distinction if I were new to [TFR]! (Wrangler, Mailing list discussion)

While wranglers might specialize in a particular kind of fandom, as fans and as TFR users their experience would likely be more reflective of the diversity of relevant fandoms. This wrangler offers two direct forms of expertise—their wrangling of video game fandoms and their experience with video games and tabletop games—and one indirect, their point of view as an imagined new user.

Similar to their expertise as users, wranglers sometimes provided evidence for guideline revision decisions based on personal experience beyond TFR. The most personal form of expertise wranglers brought to these discussions was their identity. Wranglers' personal identities—such as their ethnicities, religions, and sexual orientations—appeared as a source of expertise in discussions touching on these sensitive topics. In several instances during the "American" guideline discussion, wranglers prefaced their contributions by clarifying that they were not US-American, stating "As a non-USAn," "Answering as a citizen of English-speaking Canada," or "as a non-American of any kind." Wranglers within the United States requested this sort of expertise from their peers, recognizing the limitations of speculating from a position of privilege:

> Given how many people are reading and participating here I would like to get opinions from natives of the Americas as to how you self identify for your region. Do you prefer to see the lore of your region considered Canadian or Mexican or Brazilian regardless of your ethnicity or how much earlier than the country's founding that lore dates? Or is [sic] there ethnic or regional names that'd be preferable? […] I would like to start by getting feedback from people in parts of the Americas as to how they self identify and what their preferred identifications would be for their folklores. Please stick to your own identities so we can get an accurate picture of trends in how we might make a consistent naming pattern that can extend into the larger discussion. (Wrangler, Mailing list discussion)

This wrangler addressed her peers with respect and sensitivity, asking them to stand in as proxies for the multiple geographic identities involved in the discussion of "America."

While wranglers engaged in more formal research to determine consensus definitions, they also invited this sort of informal exchange of personal expertise when their conversations touched on sensitive identity topics represented among the wranglers.

A notable form of expertise not invoked by wranglers was that of professional background. Ongoing or past training in librarianship or archives was common among the wranglers, but wranglers did not mention these backgrounds when contributing to guideline revision discussions. Experience with traditional, established systems such the Library of Congress Subject Headings did not increase authority among the wranglers. On the contrary, in the one instance I observed a wrangler invoke such a system for emulation, the idea received sharp criticism from the wrangler's peers, who judged the Library of Congress system to be flawed and representing an undesirable US-centric and colonialist perspective.

SUMMARY

In the above sections, I described the process by which wranglers bent, violated, and rewrote TFR's internal guidelines. Given the examples I have highlighted here, it might appear as though the guidelines were constantly in flux or that they were largely ignored. This appearance is representative of how guidelines figured into complex wrangling decisions, not of wranglers' day-to-day interactions with guidelines. Wranglers most often used the examples and templates in guidelines to simplify their wrangling decisions, and wranglers typically only made small changes to guidelines, such as by providing clearer examples. Despite the diversity of the wranglers—both in terms of personal identities and in the fandoms they managed—guideline revision discussions typically resulted in consensus and agreed-upon implementations, even if doing so required an agreement to acknowledge the limited scope of those guidelines. By focusing

here on how guidelines figured into complex wrangling decisions, I revealed how wranglers understood the guidelines—as supporting wrangling work, as limited, and as built from the wrangling and personal expertise of their peers. Focusing instead on more typical interactions with the guidelines, such as consulting them for a standard form, would not have revealed this understanding.

For wranglers, the guidelines did not determine their decisions but supported their decision making. Rather than be conduits of static rules, as early classification theory defines the designer to be (Ranganathan, 1962), wranglers were representatives of their domains of responsibility—their fandoms—and of themselves as individuals of diverse identities. At TFR, wranglers used guidelines as cognitive scaffolding (Clark, 1998; Jacob, 2001), external representations of the hard work wranglers had accomplished in determining reliable—but not exhaustive—solutions to common classification design problems. Especially as they gained experience and outgrew the need for the support of this cognitive scaffolding, wranglers grew more comfortable diverging from its guidance. Experience wranglers continued to consult the guidelines to refresh their memory on the proper form for infrequently encountered tag types, but they also expressed the need to diverge from those guidelines when they contradicted a fandom's needs.

It is likely that classification designers in other contexts experience a similar relationship to guidelines in which they view them as helpful but not an absolute authority. My focus on the daily work of wranglers made visible this element of classification design, but I do not claim that it does not happen elsewhere, only that it is not prevalent in classification designers' accounts. Similarly, classification designers rarely talk about how the guidelines themselves change (for an exception, see Young & Mandelstam, 2013). Typically, narratives of change in classification design focus on how the system—its terms and term relationships—change. When classification designers

focus on how design itself changed, the change is centered as extraordinary rather than changes to design being a regular feature of a growing system (e.g., Albrechtsen & Jacob, 1998).

This analysis of how wranglers understood TFR guidelines provides insight into the role of universal guidelines in classification systems. In the following chapter, I review the three research questions with reference to the analysis in this and the previous two findings chapters.

Summary of Findings

In this chapter, I summarize the findings from the previous three findings chapters to answer my research questions. Through this chapter, I answer the research questions as follows:

- The Synthesis section summarizes the types of evidence and argument, via the factors I have identified, that classification designers used in design decisions. Together, these factors are the answer to the first research question:
  - What factors do classification designers consider in making daily decisions on terms and term relationships?
- The Interaction section examines how classification designers manage conflicts among the factors that inform their decisions. These interactions and their resolutions answer the second research question:
  - How do classification designers negotiate multiple or conflicting purposes for the classification system?
- The Guidelines section expands on the conclusion of chapter X in light of the factors to answer the third research question:
  - To what extent do rules determine the character of large scale classification systems?

In the findings chapters, I illustrated complex decisions wranglers faced and how wranglers reflected on the complicating factors. In my qualitative analysis of wranglers' complex decisions, I noted factors that wranglers identified as relevant to their decisions. Here, I shift the focus from the decisions themselves (Names, Fandoms, and Guidelines) to the factors I identified during the analysis. I first summarize the factors by defining

each factor and connecting it back to decisions in the three findings chapters. Next, I describe the factors as a set by examining how they interacted in decision, through the lens of two factors—hierarchy and autocomplete—and how they interacted with others. Finally, I reflect on the role that guidelines—wranglers' rules—played in the character of the classification system.

SYNTHESIS

In this section, I define the prominent factors that wranglers identified in their reflections on complex decisions. In doing so, I connect these emic factors to existing literature. In my analysis, I found that 10 factors were prominent in wranglers' reflections on complex decisions. I list these factors in Table 4, with "x" to indicate which findings chapters include illustrative examples of these factors. An asterisk ("*") indicates that the factor that did not figure explicitly in the complex decisions of that chapter but were relevant in other ways to that type of decision. In Table 4 and in the subsections, I list the factors in order of their prevalence in the findings chapters and in wranglers' decisions in general.

	Names	**Works**	**Guidelines**
Ambiguity	x	x	x
Filtering	x	x	x
Hierarchy	x	x	x
Temporality	x	x	x
Authenticity	x	x	x
User primacy	*	x	x

	Names	Works	Guidelines
User/designer gap	x	x	*
Inclusivity	*	x	x
Autocomplete	x		x
Server indexing burden	x		x

Table 4: Factors in complex decisions and their presence in findings chapters

I summarize the presences and absences of factors by findings chapters within the subsection below for each factor. In my summary of each factor, I limit my discussion to the factor itself, leaving discussion of interaction between the factors to the Interaction section and discussion of the factors' convergence or divergence from classification theory to the Discussion chapter. Each of the summaries is of the following structure: definition and TFR context, presence and absence in findings chapters, and context in scholarly literature. Each factor indicates a particular sense of the purpose of the classification system and a corresponding kind of evidence.

In the section summary, I answer RQ1 by categorizing these 10 factors according to four types. I present the grouping of factors according to these four types in Table 5.

Factor Type	Factors Included
Definitional	Ambiguity, Filtering, Hierarchy, Temporality
User-Centered	User Primacy, User/Designer Gap
External Truth	Authenticity, Inclusivity
Technical Context	Autocomplete, Server Indexing Burden

Table 5 Factors by factor type

The names of the four categories are etic and reflect the connections between wranglers' experience and classification theory. The names of the 10 factors are emic and reflect wranglers' descriptions of their decision-making processes. Because I relied on wranglers' descriptions to identify the factors relevant to complex decisions, these 10 factors vary in specificity or generality. That is, some factors are general concepts (e.g., "Ambiguity") while others are specific to wranglers' experience at TFR (e.g., "Server Indexing Burden").

Ambiguity

Ambiguity—the possibility for a term to refer to more than one referent—was a factor wranglers tried to remove or minimize through their design decisions. Removing ambiguity, or achieving unambiguity, was the most prevalent factor in wranglers' decisions. In fact, removing ambiguity was the primary purpose of the curated folksonomy approach: addressing the inevitable ambiguity of users' tags by creating links among tags. Wranglers expressed the goal of creating unambiguous canonical tags throughout their guidelines and training documents. Wranglers achieved unambiguity when they created canonical tags that clearly indicated a single referent. For example, "Q (James Bond)" and "Q (Star Trek)" are unambiguous tags users would immediately recognize as referring to two identically named but entirely distinct characters. Wranglers were concerned with the factor of ambiguity in two types of instances: when they sought to create a new, unambiguous tag for a referent (such as the character tag "Q (James Bond)") and when they decided how to manage user-created ambiguous tags (such as the character tag "Q").

The prevalence of ambiguity to wranglers' decisions is evident in the three findings chapters. In Names and in Works, ambiguity was a factor in every wrangler

decision. Ambiguous tags such as "Superboy" (referring to the character from *Young Justice*, a female version of that character, or a young version of Superman), created the complex problems wranglers such as Victoria tried to address. Deciding on an unambiguous canonical tag was particularly difficult in the cases of American Folklore, Bandom, and the Fiction-within-Fiction fandoms of the Works chapter. Similarly, in Guidelines, wranglers struggled to address the inherent ambiguity of "American" and debated whether "Role-playing game" and "RPF" were sufficiently unambiguous from each other. Wranglers' decisions across the three chapters indicate the complexity wranglers found in their attempts to address ambiguity.

Filtering

Filtering was the mechanism by which users used the TFR interface to compose a search query that limited results according to tags and how they further limited results by indicating additional tags on the results page. Wranglers' decisions to designate canonical tags and to create tag relationships determined the function of filtering. Specifically, canonical tags appeared to users as suggested filtering mechanisms on search results, which is why wranglers referred to tags they had not designated as canonical or as synonyms to canonicals as "Unfilterable." Furthermore, tag synonym relationships and tag hierarchies determined which tags—other than the canonical tag—would be included in filtering results. Throughout wranglers' accounts of complex decisions, they reflected on the impacts different decisions would have on users' filtering. For example, in the Names section on cape names, I illustrated how different hierarchical relationships among the superhero identities and street identities would produce different search results for users. Specifically, making Robin a subtag of both Dick Grayson and Damian Wayne would mean that users searching for Damian Wayne would receive search results with

many works featuring Dick Grayson, but which their creators had tagged with the ambiguous cape name.

Given the volume of fanworks in TFR, filtering actions were particularly important for users searching for fanworks, such as by limiting search results to fanworks featuring particular characters or themes. As evidence, in Names, Victoria discussed the filtering outcomes of different cape name structures, Brigid discussed the filtering outcomes of her team's policy for Inquisitor names, and I described the debate among Marvel wranglers over the filtering outcomes of the Daisy Johnson tag. In Works, the Bandom example and the Fiction-within-Fiction example prominently featured wrangler concerns about filtering outcomes. In Guidelines, Alba discussed the filtering outcomes of a new guideline for group and group member character tags and Catherine discussed how her fandom-specific guideline allowed for a more useful filtering mechanism in her fandom. In this manner, filtering appeared as a factor of concern in all three chapters.

Hierarchy

Hierarchy referred to the structure of subtag and metatag relationships wranglers created among canonical tags. The factor of hierarchy figured into wrangler decisions to create metatag and subtag relationships among tags, their decisions of how to order these relationships, and when to undo these relationships. The ability to create hierarchical relationships among tags was a defining characteristic of the form of curated folksonomy instantiated at TFR; in conventional folksonomies, there are no relationships among tags. Wranglers used hierarchical relationships among tags to account for differences in specificity and generality in user tag, such as by making "Alternate Universe – Modern Setting" a subtag of "Alternate Universe."

Wranglers commonly created hierarchies to instantiate relationships among tags and occasionally encountered substantial difficulty in deciding upon the best hierarchical structure. For example, in Names, hierarchy was the central concern of cape name relationships. Though wranglers found choosing which hierarchy to instantiate among related cape names and street names to be a complex decision, they appreciated the ability to create and change hierarchies as a tool to resolve the complexity of these character names. Hierarchy was also the primary means by which Brigid and her cowranglers managed the complexity of the Inquisitor character tags and by which the Marvel wrangling team managed the overlap of the Daisy Johnson character tags. In Works, the wranglers considered alternative hierarchical structures for Fiction-within-Fiction fandoms. Wranglers throughout the history of TFR, including Kerri and Dani, reflected on the hierarchical relationship between the "Bandom" fandom tag and individual band fandom tags. In Guidelines, revised hierarchy guidelines complicated Alba's management of group and group member character tags in her fandom and Catherine used hierarchical relationships to manage the centrality of original character tags to her fandom. These decisions on hierarchies touched on several other factors in wranglers' decisions; I explore the interaction among hierarchy and other factors in depth in the Interaction section.

Temporality

I use the term temporality to indicate wranglers' concerns regarding time. The factor of temporality figured into wranglers' design decisions as they considered whether current decisions would be consistent with future use. Wranglers considered current actions in terms of future use because of the intended longevity of TFR. One of the primary motivations behind creating TFR was for the fanwork community to have a

stable, central repository for its works. The fanwork community had previously lost smaller web collections as owners stopped maintaining websites or paying hosting fees and as for-profit websites abruptly changed their policies and removed substantial sub-collections without notice (Ellison, 2012). To the extent that wranglers believed in the goal of TFR as a stable repository, they put their actions in the context of the long future of TFR. Therefore, when wranglers made a decision on a particular tag, they had to consider how it was being used at the moment and how it might be used in the future. For example, each time I wrangled a player-created name—such as "Maria Shepard"—for the *Mass Effect* protagonist I considered the likelihood that, by coincidence, that exact character name would be used by the creator of a new source material work in the future, inevitability creating some confusion and the need for changes in the curated folksonomy. Wranglers' awareness of temporality also guided them to consider more stable sources of terminology, as terms they derived from these sources would remain accurate for longer. For example, choosing to disambiguate the name of a living person by their year of birth (e.g., "Jeff Davis (1975)") instead of their profession (e.g., "Jeff Davis (Producer)") was more appropriate to long-term design, as an individual might change professions over time.

Wranglers considered the factor of temporality in decisions across the three findings chapters. In Names, the factor of temporality figured into wrangler decisions in two ways—the future possibilities of user choices and the future possibilities of source material canon. With regards to user choices, Brigid and her team considered that future fanwork creators might coincidentally use the same name for a different version of their fandom's player-created character. With regards to source material changes, temporality figured into Matilde's decision for cape names in her fandom. She anticipated that, as new episodes of the source material television show aired, the cape names might be taken

on by other individuals. Similarly, Amita considered how the timing of the current release of movie adaptations for her book series fandom meant that her decisions could spoil plot developments for users who had not read the books. In Works, the possibility that users indicating different referents would use the same term came true: Kerri, Dani, and other wranglers in the history of TFR had to manage the complexity that arose from different groups using "Bandom" to refer to their fandoms. Guidelines were inherently forward-looking, as the wranglers collectively decided on how they would manage as-yet-unseen tags in the curated folksonomy. All cases in the section on making guidelines were intrinsically about the temporality of TFR. In this manner, wranglers indicated a constant concern for temporality—including whether they believed that past decisions were applicable to current tags and usage.

Authenticity

Authenticity describes the extent to which wrangler decisions were consistent with referents—such as whether character tags correctly refer to identities or whether fandom tags correctly refer to works or bodies of works. In this way, the factor of authenticity among wranglers was similar to the concept of validity in qualitative research (Creswell & Miller, 2000) and similarly was an indication of the quality of the curated folksonomy. The factor of authenticity played a role in decisions where the wranglers were concerned about how the curated folksonomy reflected an external reality, such as fanwork creators' intentions, source material canon, or cultural norms. Wranglers assessed the extent to which the curated folksonomy was authentic by how consistent its terms and term relationships were with external reality.

Wranglers reflected on authenticity in all three types of decisions I documented in the findings chapters. Specifically, in Names, authenticity was a concern for wranglers

who decided on cape name structures, as these wranglers wrestled with multiple, valid terms for character identity referents. Similarly, Brigid reflected on the limitations of authenticity in her team's Inquisitor name policy, as their decision to err towards the general would often mask the specificity of the referent the author creating a character tag had intended. In Works, authenticity was a major concern around the "American Folklore" discussion, as wranglers tried to find a term that would capture the vague yet meaningful referent. Authenticity was also a major concern for the wranglers making choices of structure in the Fiction-within-Fiction event, as they considered the real relationship between source material works and their Fiction-within-Fiction works. In Guidelines, Tamara made a principled argument on the basis of authenticity when she chose not to make canonical tags for concepts which masked the source material origin of character qualities. As a group, wranglers indicated authenticity as a priority when they sought input from experts in a phenomenon, such as members of a fan community or a geographic region. In this manner, authenticity was a factor in many design decisions.

User Primacy

User primacy referred to the idea that user choices were a fundamental and inviolable element of the curated folksonomy. Conceptually, user primacy was one of the first principles of wrangling. As a principle, user primacy dictated that wranglers respected user choices, such as by creating links among tags rather than editing user-created tags. Functionally, user primacy was always a factor since user-created tags were the subject of all wrangler decisions. User primacy was a factor each time wranglers used known user choices—such as the prevalence of similar tags—as a form of evidence in their decisions. User primacy was also a factor where user choices instigated wrangler

dilemmas, as wranglers constantly had to create new linkages to address the variation among user tagging choices.

Like ambiguity, user primacy was a factor in all wrangler decisions. Wranglers explicitly reflected on user primacy in Works, as user primacy was the defining factor in the wranglers' collective decision regarding Bandom. In that decision, wranglers chose to allow users' tagging actions to provide the de facto definition of Bandom. In Fiction-within-Fiction, wranglers encountered challenges in creating consistent disambiguation strategies because users had not been consistent in their interpretation of the ficathon's scope. Similarly, in Guidelines, Alba and Catherine argued for guideline divergence on the basis of observed user choices while Tamara reflected on her choice to defy user primacy. Wranglers who contributed to guideline revision also considered unobservable user choices, such as individuals who did not use TFR because they perceived a discriminatory bias. In this way, wranglers broadened the concept of user primacy to include non-users. Wranglers did not invoke user primacy as a factor among the examples in Names, though user choices—and the dictate that wranglers not prescribe action—were the cause of the complex wrangling tasks they described. For example, Victoria was frustrated by user choices to use cape names that she could not clearly resolve in the curated folksonomy. As a topic of explicit reflection, as in Works and Guidelines, and as an implicit root cause, as in Names, user choices were a prevalent factor in wranglers' design decisions.

User/Designer Gap

The user/designer gap was a factor in wrangler decisions that touched upon differences between the users' perspective and that of wranglers. The behind-the-scenes nature of wranglers' contributions to the curated folksonomy meant that wranglers often

considered how their decisions would appear to users and whether users would make choices consistent with those of wranglers. Key to this reflection was the wranglers' internal rule that they could not change user tags or prescribe user practice. The TFR interface instantiated this rule by limiting the users' view of tags to the tags themselves and their relationships to other tags. For example, in looking at the fandom tag "Bandom," TFR users would see the list of synonyms such as "Bandfiction" and the list of subtags, such as "My Chemical Romance," but would not see that there are 36 comments on the tag, let alone see the substance of the debate that occurred through these comments. The only way wranglers could communicate with users, within TFR, was in the creation of links between tags. Even the communication wranglers initiated outside of TFR, such as by creating "Ask the Wrangler" tumblr accounts, was subject to the policy that they not prescribe user tagging choices, so wranglers limited these tumblr accounts to updating users on new tags in the fandom and answering questions about how to perform specific searches with tags. As a result of wrangler policy and practice, users had a different view of the curated folksonomy than did wranglers and a different awareness of the effect of their own tagging choices.

Wranglers reflected on the user/designer gap in two out of the three findings chapters. Specifically, in Names, the difference between wrangler and user perspectives was apparent in the dilemmas Victoria and Catherine faced. In both cases, cape names and Inquisitor names, wranglers observed that users intended a referent with their tags—such as "Superboy" or "Ariana Trevelyan"—that the curated folksonomy system would not support. Though the users had a clear referent in mind when they tagged their works, in the broader context of the curated folksonomy, wranglers could not tie the users' tags to their intended referents. Instead, wranglers had to act in ignorance of the particulars the users intended, and linked the user tags to more general referents. Similarly, wranglers

220

anticipated that users would not understand wrangler choices. In Works, wranglers observed that their choice to use "(Fictional)" or "(Metafandom)" to disambiguate Fiction-within-Fiction fandoms would be unclear to users who were not privy to wranglers' discussions or to the wranglers' holistic view of the system. The user/designer gap was not an explicit factor in Guidelines, but this gap did influence guideline design decisions; wranglers declared that their responsibility was to interpret user intentions when creating fandom-specific guidelines and stood in as users to contribute that perspective to guideline revision discussions. In this manner, similar to the factor of user primacy, the user/designer gap was sometimes an explicit factor in wranglers' reflections and sometimes an implicit factor explaining wranglers' choices.

Inclusivity

Inclusivity refers to the potential for a classification system to represent multiple points of view and make welcome diverse users. Wranglers indicated a particular concern for how their actions impacted the inclusivity of TFR. Sensitive topics in TFR's curated folksonomy that impacted inclusivity included those related to historical and ongoing discrimination. For example, wranglers practiced extra caution when they made decisions on terms related to religion, gender, and sexuality. The wranglers practiced extra caution regarding these decisions as they recognized that their actions could favor one group over the other or make some potential users feel unwelcome at TFR. For example, designating some supernatural traditions as "religions" and others as "mythology" could have indicated greater respect for written traditions (e.g., "Abrahamic Religions") than for oral traditions (e.g., "Native American Mythology").

Inclusivity appeared in findings chapters where wranglers considered how their decisions could ameliorate or reinforce historical discrimination. In Works, wranglers

who sought an alternative to the term "American Folklore" shared a concern to treat all sacred traditions—religions and folklore traditions included—with equal respect. At the same time, the wranglers were aware of a tendency for TFR to instantiate US-centric views and language, and they struggled to find a solution that would not reinforce this bias. To a lesser extent, wranglers' discussions on Bandom and Fiction-within-Fiction terminology reflected wranglers' concern not to practice discrimination by siding with one group or another. In Guidelines, wranglers revisited the same concern regarding a US-centric bias in their discussion of the term "American." I do not include Names as a chapter which featured sensitive topics as a factor. The wranglers did not discuss complex decisions which touched on sensitive topics; the most prominent sensitive topic in decisions on character tags was how wranglers handled canonical tags for transgender characters. This decision was not complex because the wranglers had already decided the issue conclusively; regardless of all other factors, a transgender character would always be known by the character's chosen name. This absence highlights the blind spot inherent to my focus on complex decisions: my analysis does not feature factors that were so important to wranglers in classification decisions as to be beyond debate. Therefore, in the window of time in which I studied TFR, the factor of inclusivity was more salient where wranglers had not yet settled such issues—as in Works and Guidelines—and less salient where the wranglers had settled the issue definitively—as in Names.

Autocomplete

Autocomplete is a common factor in many contemporary user interfaces in which the system offers suggestions to complete a user's entry to a text field. The system typically makes autocomplete suggestions by matching the characters the user is typing into a text field against either a computed prediction (developed from a corpus of past

entries) or a designer's finite list (as in the case of TFR) (Cai & de Rijke, 2016). In TFR's instantiation of autocomplete, canonical tags would appear to users as suggestions in a drop-down list as the user entered similar character strings into tag fields. Each time wranglers made a decision about a new canonical tag, and therefore a change to the finite list of autocomplete suggestions, they considered how their choice would affect autocomplete and how autocomplete would in turn affect the user's experience of the curated folksonomy. Autocomplete was therefore a factor in many wrangler decisions regarding the creation of canonical tags.

Autocomplete appeared as an explicit factor in wrangler decisions in two of the three findings chapters. In Names, Matilde and the Marvel wrangling team considered how intervening early to create a canonical name for a character would influence future users—through autocomplete suggestions—to more consistently use an unambiguous form of the character's name. Conversely, Amita made a choice to keep some information out of a character's canonical tag to avoid spoiling users who had not finished the book series. In Guidelines, Tamara avoided making a canonical tag because she believed that its presence in autocomplete would reinforce a type of tag she thought was not in keeping with the relationship between canon and fanworks. Autocomplete did not appear as an explicit factor in decisions in the Works chapter. It was, however, implicit in wranglers' reflections on the "Bandom" fandom tag and the possible "Tall Tales" fandom tag. In both these cases, wranglers were concerned that users could apply the fandom tags more widely than the wranglers had intended. In part, they anticipated users adopting a tag through its appearance in autocomplete; autocomplete made fandom tags visible but stripped of context, so wranglers anticipated how users might interpret fandom tags without context. In this manner, wranglers' awareness of the function of

autocomplete meant that it was often a factor in their design decisions regarding canonical tags.

Server Indexing Burden

Wranglers used the term "server indexing burden" to refer to the number of changes that wrangling decisions required at the level of TFR's web server. The server indexing burden was a factor in wrangler decisions that would noticeably impact the function of TFR by triggering a substantial number of indexing changes in the curated folksonomy. The server indexing burden became a central concern among the wranglers as users—including wranglers—experienced noticeable latency and frequent failures when navigating TFR. After examining the causes of this phenomenon, the TFR technical development team concluded that some wrangling actions triggered so many changes to the server's index of works that the server's time resolving them negatively affected the website's response time. As the TFR technical development team and wranglers came to better understand the relationship between wrangling actions and the server indexing burden, the two teams agreed on approaches to minimize the impact. These approaches included periodically "turning off the wrangulator"—suspending access to the wrangling interface—for hours at a time to a day per week to allow the server to "catch up;" scheduling changes to large tags—those linked to more than 5000 works—to low traffic times of day; and minimizing deep tag hierarchies.

Wranglers considered the server indexing burden as a factor in their decisions in two out of the three findings chapters. In Names, awareness of the server indexing burden motivated the Marvel wranglers to resolve the ambiguity of the "Daisy Johnson" tag as soon as possible, as they anticipated that the growth in use of the tag would soon reach over 5000 uses, necessitating changes then be scheduled with and routed through staff

wranglers. Similarly, Justine considered the timing of changes to character names in her fandom to minimize the number of times tags would have to be changed. Since relationship tags (e.g., "Ned Stark/Cathryn Tully") contain two or more character tags, she made efforts to make simultaneous changes to character tags in her fandom, so that their relationship tags would only have to be changed once, making fewer indexing requests on the server. The relationship between the server indexing burden and hierarchy prompted the discussions in Guidelines. I explain this relationship in detail in the Hierarchy subsection within the Interaction section, as the server indexing burden came to affect how wranglers made decisions on hierarchies. The server indexing burden did not appear in wranglers' discussions of Works. This absence is partially due to the distinct function of fandom tags; wranglers needed to decide upon a fandom tag at the outset of fanwork activity for the fandom since fandom tags determined disambiguation for other tags and organized wrangler tasks. Therefore, it was less common that wranglers made changes to fandom tags linked to more than 5000 works, triggering a substantial server indexing burden. In this manner, the server indexing burden was particularly salient as a factor in maintenance decisions—decisions responding to constant changes in the curated folksonomy—and less salient in initial design decisions—decisions wranglers made ahead of trends in user choices.

Synthesis Summary

The factors in this section were important to wranglers' complex design decisions. Above, I ordered the factors according to their prevalence in the findings chapters and, within the section for each factor, I noted ways in which they figured into decisions outside of that analysis. This list of factors is a partial, literal answer to the RQ1: "What factors do classification designers consider in making daily decisions on terms and term

relationships?" Here, I categorize the 10 observed factors to answer a more abstract sense of RQ1. That is, "What types of factors do classification designers consider in making daily decisions on terms and term relationships?" Here, I categorize the 10 factors within four types of factors: definitional factors, user-centered factors, external truth factors, and technical context factors.

Definitional factors

The first type of factors wranglers considered were those that define classification systems: ambiguity, filtering, hierarchy, and temporality. Each of these factors is necessary for my broad definition of classification systems: "systems of terms and term relationships intended to sort and gather like concepts and documents." The factor of ambiguity is implicit in the use of "terms" and the concept of "likeness." Together these elements of the definition refer to the notion of setting boundaries between what is sufficiently alike and what is different by concretizing referents in unambiguous terms. Filtering is another term for the phrase "sort and gather," the key functions enabled by classification systems. Hierarchy is one type of "term relationship" among other possible term relationships, such as related term relationships, that define a classification system. The final term, temporality, is implicit but no less important. As classification systems are designed systems necessitating substantial effort and collecting a set of documents or concepts, we assume these systems are meant to operate over time. This idea of temporality is particularly evident in Buckland's (Buckland, 2012) notion of obsolescence—the phenomenon by which the meanings in descriptors drift away from their referents over time—which I explore in more depth in the Discussion chapter. The prevalence of these four factors to wranglers' design decisions is therefore self-evident;

to the extent that the wranglers were classification designers, the factors of ambiguity, filtering, hierarchy, and temporality figured prominently in their decisions.

User-centered factors

The second type of factor wranglers considered in their design decisions was specific to a particular school of classification theory: user-centeredness. In the literature review, I showed how scholars have presented user warrant as a method for designing classification systems that embody user terminology and therefore adapt to the users' needs. The factors of user primacy and the user/designer gap represent the wranglers' version of this school of classification theory. The factor of user primacy was an extreme version of user warrant. This form of user warrant was embodied through the curated folksonomy approach in which user tagging choices were the raw material of classification design and through the principle by which wranglers limited their actions to making links rather than changing users' tags. Similarly, when wranglers considered the user/designer gap as a factor and made efforts to adapt their system to the users' perceptions, they privileged the user-centered view typically embodied by the application of user warrant. Through consideration of these two factors, user primacy and the user/designer gap, the wranglers' approach to classification system design converged with that of user-centered designers in classification theory.

External truth factors

The third type of factor stood in contrast to the second factor in that this type of factor indicated the wranglers' sense that the classification system had a larger function beyond connecting users to documents. Wranglers considered the factors of authenticity and inclusivity in their decisions in order to make the classification system represent an external truth. I observed a parallel contrast in my literature review between literary or

scientific warrant and user warrant. Both instances contrast a view that the relationship between terms and referents is entirely internal to the system—that the users recognize the terms as applying to the documents they find relevant—against a view that the relationship between terms and referents was to be made with reference to an external authority—such as a domain of literature or the consensus of experts. To extend the contrast further, the user-centered factors represented the world as users understood it, the factor of authenticity represented the world as it was, and the factor of inclusivity represented the world as it should be. By considering external truth factors, wranglers indicated that they believed that the classification system served a purpose other than the instrumental filtering function. That is, they felt responsible to create a system that represented an accurate and respectful vision of the world.

Technical context factors

The fourth and final type of factor wranglers considered in their decisions was the acknowledgement of the technical context and entanglements of the classification system. Wranglers considered the factors of autocomplete and the server indexing burden because they understood that their design decisions had effects on users beyond effects on the users' ability to find relevant fanworks. These factors have no parallel in schools of classification theory I represented in my literature review. Instead, the presence of these factors in wranglers' design decisions represented an infrastructural view of classification systems which included additional layers of infrastructure—the user interface and the web server. By considering these factors, the wranglers indicated that decisions which provided more effective filtering, aligned the system to users' terminology, or aligned the system to an external truth, were not sustainable if those decisions negatively impacted

the function of the site itself, either by causing future ambiguity through autocomplete or making the site unusable through too great server indexing burden.

Together, these four factors—definitional factors, user-centered factors, external truth factors, and technical context factors—represent the considerations wranglers made in their daily design decisions. In the Discussion chapter, I explore each of the factors in more depth and identify the implications that these four types co-existed within the wranglers' design paradigm. In the following section of this chapter, I discuss this co-existence in another way, through analysis of how these factors overlapped and contradicted each other.

INTERACTION

In the previous section, I summarized each of the 10 factors as if they were isolated and independent factors in wranglers' classification decisions. These summaries served to explain each of the 10 factors but did not account for how the factors overlapped and conflicted. My second research question asked how designers resolve conflicting responsibilities, which I have operationalized here as the factors they considered in their design decisions. In this section I focus on wranglers' considerations of overlapping and conflicting factors to illustrate how wranglers' understood their multiple responsibilities. For example, wranglers' considerations regarding the fandom tag "Bandom" involved weighing whether wranglers should facilitate effective filtering or accept all user choices as valid. In the first option, the wranglers would see their primary responsibility as creating an effective controlled vocabulary—a dominant idea in classification theory—or represent the cacophony of voices in the TFR community—an idea represented in critical, contemporary classification theory.

To so illustrate how wranglers resolved the multiplicity of factors relevant to their decisions, I focus on two factors—Hierarchy and Autocomplete—and document how they overlap and conflict with others. Through these two factors, I account for interactions with each of the other eight factors and illustrate instances in which wranglers judged one factor to be more important than the others. In the section summary, I return to RQ2 and address how classification designers in this community resolved their conflicting responsibilities.

Hierarchy

I noted in the section on hierarchy, above, that this factor is foundational to a variety of classification systems. It should not then be surprising that this factor intersects with many others in classification designers' decisions. In this section on interaction with hierarchy, I document how filtering, authenticity, user primacy, and the server indexing burden intersected with this factor, and which factors were most important to wranglers' decisions.

Wranglers used hierarchy to facilitate more complex filtering choices for users. For example, a hierarchy in which "Sherlock Holmes & Related Works" was a broader term to "Sherlock Holmes – Arthur Conan Doyle" and "Elementary (TV)" gave users the opportunity to filter works by one of the latter two versions of Sherlock Holmes or to return works from both sets, if the user was interested in all versions of Sherlock Holmes. Despite this clear relationship between hierarchy and filtering, the factors are not identical; wranglers also considered authenticity in creating hierarchical relationships among terms. Wranglers predominately chose to place Fiction-within-Fiction fandoms below their source fandoms in hierarchy—rather than parallel to them. This choice sacrificed utility in filtering for authenticity in representing the relationships among

fandoms; users could not limit their searches to the broader term, the source fandoms. Hierarchy could facilitate filtering and it could represent authentic relationships; when it could not do both, wranglers chose authenticity.

The intersection of hierarchy and authenticity drew in the factor of user primacy in the case of "Bandom." The wranglers' choice to allow user tagging actions to define Bandom created a hierarchy in which over a hundred band fandom tags were direct subtags to the Bandom fandom tag. Bandom was such a general term as to potentially refer to any one of a number of sets of band fan communities; rather than adjudicate which fan community had the most authentic claim to the term and police the hierarchy accordingly, wranglers chose to follow the factor of user primacy. Wranglers removed themselves from the debate over authenticity, sacrificed precision filtering, and allowed the users themselves to determine the scope of a term upon which they, the users, had historically been unable to agree. In classification design, creating hierarchy involves making implicit statements on what things are and where they belong; creating a hierarchy in which "My Chemical Romance" was below "Bandom" implicitly stated that My Chemical Romance belonged to Bandom, that it *was a* Bandom. Though the obvious impact of hierarchy was its effect on filtering, hierarchies also unavoidably made claims on the authenticity of the relationships they instantiated. Determining hierarchy by user primacy, as a practice, made clear that users were the authoritative voice regarding what relationships were authentic.

The potential for hierarchy to facilitate filtering and instantiate authentic relationships tempted wranglers to create hierarchies throughout the curated folksonomy, some with many layers of broader-narrower term relationships. This utility, however, conflicted with the overall function of the site as tags in hierarchical relationships placed a disproportionate strain on the server. As I explained in the Guidelines chapter and for

231

the server indexing burden factor, tags in hierarchies created a disproportionate demand on the server to index connections between tags and works. For wranglers, this demand meant that new hierarchies had to make more substantial contributions to filtering for wranglers to justify the additional server indexing burden. The discovery of the server indexing burden also necessitated that wranglers review existing hierarchies to identify which should be undone, a process wranglers referred to as "tree trimming." This technical factor, one that became visible through the quick rise in use of TFR, became an important catalyst for the guideline revision process I described in the Guidelines chapter, as wranglers recognized the extent to which they would have to change their practices to minimize existing and new hierarchies.

Autocomplete

Like the server indexing burden, autocomplete was a technical factor that wranglers considered as they observed the connections between their wrangling decisions and the users' experience of the site. As I noted above, wranglers' reflections on autocomplete indicated more consideration for the reciprocal effect of design on user language than is typical of design literature in this area. In this subsection, I will describe how wranglers understood autocomplete's effect in light of its interactions with other design decision factors, including user primacy, authenticity, temporality, and sensitive topics.

Wranglers often made arguments from user primacy based on the number of uses of a given tag. For example, Amita cited as evidence for keeping an incomplete character name as a canonical tag the overwhelming usage of the incomplete version among users. A staff member reminded Amita that "format does influence" usage. The staff member here was indirectly referring to the function of autocomplete as influencing user tagging

choices; though users may have initially intended to type out the character's complete name, entering the first name would have triggered a suggestion to use the existing canonical. Agreeing to this suggestion—by pressing "Enter" rather than finishing typing the complete name—would be less effort. Autocomplete may have also influenced user tagging choices by implicitly identifying the incomplete character name as correct. Given the care wranglers took in choosing which tags to make canonical, autocomplete suggestions were more consistently formatted and freer of grammatical errors than other tags in the site; users would reasonably expect autocomplete suggestions to be the "right" tag and would become used to agreeing with autocomplete's suggestions—the assumption that autocomplete suggestions were authentic granted them authority. Similarly, autocomplete would influence user tagging choices in those cases where the user understood the function of the curated folksonomy well enough to know that autocomplete suggestions were canonical tags. Choosing to type a tag other than the autocomplete suggestion might mean that the user's tag—and the user's associated fanwork—would remain less visible in search until a wrangler linked the user's tag to the existing canonical. Users who understood the function of the curated folksonomy and wanted their fanworks to be visible to other users would be motivated to agree to autocomplete's suggestions. All of these user choices would increase the tagging counts that wranglers such as Amita relied on as indicators of user primacy, blurring the line between wrangler influence and user choice.

Given the potential impact of autocomplete on subsequent user tagging choices, the intersection of autocomplete with the factor of temporality was inevitable. The TFR interface used autocomplete not only to save users' time in typing in tags; influencing users to choose existing tags made the curated folksonomy more efficient as future users

were more likely to converge on a single term[7]. Therefore, the factor of autocomplete always intersected with temporality, as wranglers considered tags in autocomplete would influence user choices and would change the set of unwrangled tags the wranglers had to manage. Specifically, wranglers used autocomplete to influence users to choose unambiguous tags, such as when Matilde hoped to influence users to choose the street identity name for a superhero character.

Autocomplete intersected as well with the factor of temporality when wranglers considered when a given autocomplete suggestion would be appropriate; Amita agreed that the complete name of the character was right but considered that the ongoing release of movie adaptations meant that the complete name would spoil plot points unknown to many of the current users of TFR. In light of autocomplete, a wrangler's choice may have been correct at a later point in time but harmful in the present. In cases such as Amita's consideration of spoilers, the concern was not which tags autocomplete influenced a user to choose, but what undesirable information autocomplete conveyed to the user.

Similarly, autocomplete intersected with wranglers' concern for authenticity and sensitive topics in the established policy regarding transgender names. A wrangler could make an argument based on authenticity to include a transgender character's former name—known as their "dead name"—within a canonical character tag. This argument would be that, at points in the source material's canon, the dead name was the character's correct name—it would have been used in the source material and in metadata such as cast listings. For a user with an incomplete knowledge of the canon, the dead name might be the only one they know, and including that name in autocomplete—such as by making "Unique 'Wade' Adams" the canonical character tag for a character with the dead name

---

[7] This design strategy was also used at Delicious, where the interface listed common tags for a given link, speeding the typical process by which user choices would converge on a small set of tags (S. a. Golder & Huberman, 2006; Munk & Mørk, 2007).

Wade—would help that user find relevant fanworks. However, this canonical tag would also appear in autocomplete for users entering "Unique Adams." Presenting those users with the character's dead name and implicitly legitimizing that dead name as authentic would be seen as disrespectful if not violent (Gratton, 2016; Haimson, Brubaker, Dombrowski, & Hayes, 2016). Like Amita's consideration of spoilers, the wrangler policy not to include dead names in canonical character tags indicated a concern for what autocomplete should not have shown out of respect, rather than what it should have shown for utility.

As these intersections show, wranglers' awareness of autocomplete as a factor in their design decisions reinforced some factors and overrode others. Though autocomplete facilitated wranglers' desire to minimize ambiguity, by influencing users to choose the unambiguous tags wranglers had designated as canonical, it also muddied interactions of user primacy and could have presented users with undesirable information. Consistent with how wranglers resolved the interactions among hierarchy, authenticity, and filtering above, when wranglers faced complex wrangling decisions involving autocomplete they ultimately made their choices based on how they believed the site should appear to users rather than how well it might function as a retrieval mechanism.

Interaction Summary

To answer RQ2, on how classification designers negotiate multiple or conflicting responsibilities, one answer from TFR is that designers in a single system—or even a single designer at different times—will adjudicate conflicting factors in different ways. For example, wranglers will sometimes prioritize filtering and sometimes authenticity. At TFR, wranglers typically made choices in complex decisions that favor the responsibility of the system to represent the values of the community rather than the function of the

system in facilitating retrieval. Factors such as authenticity and inclusivity repeatedly overturned factors such as filtering. In prioritizing the factors in this way, wranglers embodied a philosophy of classification design more consistent with contemporary, critical approaches (e.g., Albrechtsen & Jacob, 1998; Feinberg, 2011) than with the objectivist foundations of classification theory (e.g., Bliss, 1929; Ranganathan, 1962).

Though the definitional factors were most prevalent in complex decisions, and the user-centered factors most prominent among the wranglers' principles, external truth factors and technical context factors were often sufficient to overturn these other considerations. The reason that the latter two types of factors were less prevalent may be that wranglers only needed to explicitly consider external truth or technical context when the definitional factors and the user-centered factors would lead to harm. In this sense, the former two types of factors were foundational, and supplied wranglers with the necessary information to make most decisions, while the latter two types of factors represented greater responsibilities upon which the wranglers would not infringe. As an example of this interaction among the types of factors, I return to the example of the wranglers' policy for transgender character names. There, matters of filtering and user primacy— where it may have been more efficient for wranglers to include the characters' dead name in canonical character tags—were less important than inclusivity to LGBTQ users and authenticity to an individual's sense of self. In this case, the design decision had ceased to be complex once the wranglers established the latter factors; the greater responsibilities to create an accurate and respectful system eventually became embedded in practices and wranglers no longer explicitly reflected on them.

In the context of TFR, wranglers resolved conflicting responsibilities by erring on the side of their responsibility to create an accurate and fair classification system. While classification designers in other settings would likely arrive at other prioritizations, here

the prioritization of external truth factors above definitional or user-centered factors is consistent with the community and its collection. I explore the implications of the fit between the classification and the TFR community in depth in the Discussion chapter. In the next section of this chapter, I shift to the discussion of wranglers' guidelines. While this section has revealed how wranglers resolved conflicts among factors in practice, the next addresses how wranglers sought to embody their priorities in guidelines for future decisions and how wranglers understood guidelines as being sufficient—or not—to determining future decisions.

GUIDELINES

In the third findings chapter, Guidelines, I detailed how wranglers diverged from their guidelines, justified these divergences, and developed new and revised guidelines for future decisions. Through that analysis, I illustrated what wranglers meant by "wrangler discretion." The idea of wrangler discretion placed a wrangler's expertise with regards to their own fandom above the authority of TFR-wide guidelines. Wranglers were familiar with the "weirdness" of different fandoms such that they would justify divergences from TFR-wide guidelines by standing as experts on source material, tagging practices, and user needs.

For wranglers, the guidelines did not determine their decisions but supported their decision making. Rather than be conduits of static rules, as early classification theory defines the designer to be (Ranganathan, 1962), wranglers were representatives of their domains of responsibility—their fandoms—and of themselves as individuals of diverse identities. At TFR, wranglers used guidelines as cognitive scaffolding (Clark, 1998; Jacob, 2001), external representations of the hard work wranglers had accomplished in determining reliable—but not exhaustive—solutions to common classification design

problems. Especially as they gained experience and outgrew the need for the support of this cognitive scaffolding, wranglers grew more comfortable diverging from its guidance. Experienced wranglers continued to consult the guidelines to refresh their memory on the proper form for infrequently encountered tag types, but they also expressed the need to diverge from those guidelines when the guidelines contradicted their fandoms' needs.

It is likely that classification designers in other contexts experience a similar relationship to guidelines in which designers view them as helpful but not an absolute authority. My focus on the daily work of wranglers made visible this element of classification design, but I do not claim that it does not happen elsewhere, only that it is not prevalent in classification designers' accounts. Similarly, classification designers rarely talk about how the guidelines themselves change (for an exception, see Young and Mandelstam, 2013). Typically, narratives of change in classification design focus on how the system—its terms and term relationships—change. When classification designers focus on how design itself changes, the change is presented as extraordinary rather than change being presented as a regular feature of a growing system (e.g., Albrechtsen & Jacob, 1998).

This analysis of how wranglers understood TFR guidelines provides insight into the role of universal guidelines in classification systems. In my third research question, I asked: "To what extent are principles or guidelines appropriate for large scale classification systems?" The answer, from the case of TFR wranglers, is that guidelines support but do not determine the ongoing design of large-scale classification systems. Guidelines were resources for the wranglers, similar in function to user-generated resources such as fandom wikis. Wranglers typically found guidelines to provide clear solutions to infrequently encountered tag problems. Wranglers did not follow guidelines when guidelines contradicted the wranglers' judgment for a specific case. Indeed, many

of the complex design decisions through the three findings chapters detail wranglers' experiences that the guidelines suggested a design decision that would run counter to one or more of the factors I identified.

The process by which wranglers developed and revised guidelines made clear the wranglers' collective values regarding the purpose of the classification system and the limitations of universal guidelines. Therefore, in addition to providing support for daily design decisions, guidelines also served as representations of the designers' common understanding. The acts of questioning and revising the guidelines reinforced the wranglers' values and provided periodic opportunities to align the classification system to the wranglers' basic principles. A second answer to RQ3, then, is that—aside from the function guidelines play in supporting or determining design decisions—the process of guideline development itself develops and reinforces designers' values. Guidelines embodied the consensus of wranglers regarding a number of design decisions, enabling wranglers to create fandom-specific guidelines consistent with their principles. Therefore, guidelines are appropriate to large-scale systems in that they articulate designers' values back to the designers.

SUMMARY

In this chapter, I have answered my three research questions using illustrative examples from the previous three findings chapters. I found that wranglers considered 10 factors in making classification design decisions, belonging to four types: definitional factors, user-centered factors, external truth factors, and technical context factors. When resolving conflicts among those factors, wranglers considered whether decisions based on definitional and user-centered factors violated external truth and technical context factors. While the former two types of factors were more prevalent in design decisions, wranglers

prioritized the latter two types to resolve conflicts. Wranglers' prioritization of external truth and technical context factors indicated that, while they ascribed to a user-centered classification design approach, wranglers ultimately believed that this approach was secondary to their responsibility to create an accurate and respectful classification system.

Wranglers strove to create guidelines that would achieve consistency across TFR and that would support wranglers' decision making. In doing so, wranglers reflected on the purposes of the classification system, demonstrated the types of evidence they believed were valid inputs to universal guidelines, and collectively established guidelines as useful but limited. True to the term wranglers chose, "guidelines," wranglers' rules and standards provided guidance without determining their decisions.

In the next chapter, Discussion, I explore the implications of the above findings and turn to the pressing problem of the relationship between classification design and the character of classification systems. To do so, I return to the literature on warrants and the skills of classification design, noting where TFR wranglers enacted established principles of classification design and where they diverged.

Discussion

In the findings chapters, I focused on how classification designers at TFR made design decisions and the factors they considered in those decisions. In this chapter, I place those decisions in the context of classification theory. In some respects, the approach of classification designers at TFR converged on established approaches from classification theory. These instances of convergence help to illustrate classification theory principles. In other respects, classification designers' experiences at TFR were not well represented by classification theory. I note such instances of divergence to illustrate the role of designers' agency and the context of classification decisions in creating the character of classification systems.

In the first section of this chapter, I review each of the 10 factors with references to classification research literature and infrastructure studies literature. Through the 10 factors, I bring together these two literatures, revealing their shared and different perspectives on ubiquitous issues in classification design. I then turn to issues of generalizability and specificity to explore how my findings at TFR are indicative of the designers' role in other settings.

TFR FACTORS IN THE LITERATURE

In the following subsections, I review each of 10 factors I identified as TFR designers' considerations in design decisions. Whereas in the previous findings chapter I reviewed the instances of each factor and its role in design decisions, here I compare the emic factors at TFR to principles from classification theory and infrastructure studies, such as warrants, skills, and embeddedness.

Ambiguity

Just as reducing ambiguity was a prevalent factor in wranglers' decisions, ambiguity has been a major theme in classification theory. Classification theory is about the inherent ambiguity of natural language (Szostak, 2008b) and its solution in classification design (Mayor, 2012; Mills, 2004; Tudhope & Binding, 2008). In classification theory, reducing ambiguity has typically been presented as one of the primary purposes of classification design; Zeng (2008) listed "eliminating ambiguity" as the first of the four fundamental functions of classification systems. Conversely, other classification scholars have challenged the primacy of unambiguity as a goal of classification systems. This challenge is sometimes on the basis that ambiguity is the natural state of the world, and that unambiguity is therefore inaccurate (Feinberg et al., 2014a) or that embracing ambiguity allows designers to make more creative and useful decisions (Theimer, 2012).

Despite the centrality of reducing ambiguity to the curated folksonomy, I observed cases in which wranglers strategically deployed ambiguity. The Bandom example, in particular, revealed that wranglers sometimes used ambiguity to avoid the harm that unambiguity would cause. In such instances, wranglers displayed an attitude towards unambiguity that was more consistent with the contemporary challenge in classification theory to embrace ambiguity (e.g. Feinberg et al., 2014), though their straightforward decisions still reflected the more typical dedication of classification designers to reduce ambiguity. Olson (2001a) predicted this strategy to preserve some ambiguity to avoid imposing one point of view. In her critique of sameness and difference Olson noted that classification designers make judgments on which sorts of meanings are most salient and advised that designers question whose meaning they privilege through their decisions on terms and term relationships. In that critique, Olson

emphasized the importance of local and multiple solutions to classification problems over "one-size-fits-all classification" (p. 121); accepting ambiguity was a divergence from the "one-size-fits-all" approach of using disambiguated terms and was a strategy wranglers employed sparingly and after a great deal of reflection.

Filtering

Filtering was an emic term that arose among the wranglers, but it has clear synonyms in classification theory, where scholars describe the factor of filtering using the terms "precision" and "recall" (Bates, 1986; Cleverdon & Keen, 1966; Golub, 2006; Ibekwe-SanJuan, 2006; Markey, 2006; van Rijsbergen, 1979). Precision and recall are common operationalizations of the concept of filtering; "relevance" from information retrieval research is another (e.g., Zhang, Zhang, Lease, & Gwizdka, 2014). Precision is a measure of how few irrelevant items a search returns; the fewer false positives (results the system incorrectly interpreted as relevant), the greater the system's precision. Recall is a measure of how many of the relevant items a search returns; the fewer false negatives (results the system incorrectly interpreted as irrelevant), the greater the system's recall. For any retrieval-oriented system, precision and recall are relevant concerns, regardless of the philosophy of classification the designer follows. The importance of precision and recall holds true for systems from library classification schemes to web search algorithms (Bates, 1986; Golub, 2006).

Though wranglers did not use these terms, they carefully considered precision, recall, and relevance effects when they considered the effects of their wrangling decisions on user filtering, mirroring the concerns of classification theory and information retrieval research. The difference between the approach of wranglers and that of classification designers in the literature is that the former considered the measures informally and from

their dual perspective as users. Rather than test precision and recall as objective metrics (as in Ibekwe-SanJuan, 2006), wranglers considered filtering conceptually and only occasionally referred to numbers from search results when debating filtering effects. Without access to the research labor and measurement tools scholars in classification theory and information retrieval use to assess filtering, wranglers instead appealed to notions of aboutness (Hutchins, 1978), the conceptual underpinning of all these concepts. For example, instead of having expert evaluators (Razikin, Goh, Chua, & Lee, 2011) or crowd workers (Zhang et al., 2014) judge the number of search results that matched a query, wranglers considered whether fanworks *about* One Direction would also be *about* Bandom. For wranglers, therefore, determining filtering meant considering factors regarding aboutness and the users' perspectives—authenticity and user primacy—rather than an independent measurement such as precision.

Hierarchy

In classification theory, as in TFR's curated folksonomy, hierarchy has been a common tool to resolve specificity and generality among concepts in a collection. Hierarchy is a common feature of many knowledge organization genres, such as thesauri, taxonomies, ontologies, and library classification schemes (Aitchison et al., 2000; Gilchrist, 2003). Among knowledge organization genres, taxonomies make the potential of hierarchy most apparent; taxonomies instantiate levels of hierarchy to facilitate information organization and browsing at multiple levels of specificity (Gilchrist, 2003). One of the most immediately familiar taxonomies, the Linnaean taxonomy, sought to organize all living organisms within a single hierarchical structure, reflecting organisms' similarities and shared evolutionary histories (Ereshefsky, 2000; Knorr-Cetina, 1999; Mayor, 2012). Hierarchies help to order a universe of knowledge. Therefore, one

prominent criticism of these approaches is that classification designers impose this order at the expense of masking or manipulating more complex relationships (Olson, 2004). Within TFR, experienced wranglers echoed Olson's criticism when they cautioned their peers not to create order for the sake of order.

Difficulties that the wranglers encountered in instantiating hierarchies occurred because of the interaction between hierarchy and filtering. That particular mode of interaction—that lower levels of the hierarchy were always returned on searches for the higher levels—was a design choice made before the conceptual work of the classification designers and was beyond their agency to alter. Whereas the above literature has discussed hierarchy as more or less conceptually correct, hierarchy can also be judged according to its operation within a system, as wranglers considered it.

Temporality

Temporality has long been an implicit factor in classification theory. Buckland made the factor of temporality explicit in his relatively recent discussion of obsolescence (Buckland, 2012). In that article, Buckland identified classification design's dual nature as forward-looking—thinking about the information needs of future users—and backward-looking—built from the extant collection and reflecting the discourse scholars have established over decades of work. This dual nature occurs because classification systems are an inflection point between accumulated knowledge and future use. Wranglers' experiences were consistent with this factor of classification design, and their decisions reflected a concern for the immediate future of the system (e.g., how well their decisions would support users finding a story using a tag under review) and the long future of the system (e.g., how consistent their decisions would be with the growth of source material and future user tagging choices).

While the notion of temporality in classification theory is a latent and occasionally explicit factor in research, infrastructure studies have consistently focused our attention to the temporality of infrastructures. In particular, infrastructure studies' focus on maintenance has revealed how much of the character and function of an infrastructural system is attributable to choices in maintenance work (Lee, Dourish, & Mark, 2006; Ribes & Bowker, 2009; Ribes & Lee, 2010; Star & Ruhleder, 1996). Because we build infrastructural systems—such as electrical grids and highways—to last for a long time and to be the backbone upon which we build not-yet-imagined infrastructures (e.g., fiber optic internet and autonomous, networked trucks), temporality is a prominent factor in their design, even as the passing of time makes once spectacular systems fade from view. Wranglers' reflections on the relevance of temporality to their design decisions were consistent with this view and linked the latent concern over temporality from classification theory to the foregrounded concern over temporality infrastructure studies. More than other factors, temporality makes clear that imagination and creativity are core skills in classification design; it is the designers' responsibility to think beyond the current collection to how their choices will align with unknown items, users, and topics.

Authenticity

Authenticity as a factor in classification design decisions appears in classification theory through appeals to literary warrant and scientific or consensus warrant. Both these forms of warrant help classification designers make decisions on the basis of an external reality. Decisions based on literary warrant are meant to reflect authorial intentions and the discourse of a scholarly community (Clare Beghtol, 1986, 1995; Hjørland & Albrechtsen, 1995). Decisions based on consensus warrant are meant to reflect reality—

246

whether meaning scientific reality or social reality—as it has been determined by groups the designer recognizes as cognitive authorities (Clare Beghtol, 1986; Bliss, 1929; Rieh, 2005; Wilson, 1983). These forms of warrant are classification designers' means of capturing reality, of making the classification system an authentic representation of knowledge.

By soliciting expert input and attempting to make terms good representations of referents, wranglers indicated a concern with authenticity consistent with traditional approaches to classification design. However, authenticity also served a more practical purpose than the principle goal of aligning the classification system with an external reality. With reference to the factor of temporality, classification designers recognized that terms based on scientific or consensus warrant were more likely to be stable for a longer period of time. That is, in order to facilitate the longevity of TFR's relevance to users, it was prudent for wranglers to draw from sources of terminology that were not subject to the same trends and fads as user terminology. To the extent that wranglers intended for the classification system to serve users beyond those currently represented through their user-created tags, wranglers chose sources of terminology belonging to a broader context than TFR itself—namely, the source material works themselves and agreed upon terminology for marginalized populations.

User Primacy

Similar to user-centered design in human-computer interaction (Gould & Lewis, 1985; Norman & Draper, 1986), designers who centralize user primacy in classification design create design requirements from and evaluate design with user data. User primacy has been represented in classification theory through the application of user warrant (Bates, 1986; Clare Beghtol, 2002; Mayor, 2012; Rafferty & Hidderley, 2007; Wang et

al., 2008). Though designers always make classification systems with use in mind (aside from a few exceptions such as demonstrations or thought experiments, e.g. Feinberg, Carter, & Bullard, 2014b), a designer's choice to make user warrant the primary basis for a system indicates a more radical alignment with user needs than would the choice to follow literary warrant. The main difference in serving users through user warrant or literary warrant is whether the designer believes that users are qualified to determine the best language for the system (Clare Beghtol, 1986; Hjørland, 2013b).

The curated folksonomy at TFR complicated the typical division between user warrant and literary warrant since the collection—the literature to which designers typically turn to generate literary warrant—was itself user-generated. The nature of TFR as a user-created, user-run website collecting user-generated works made the choice to prioritize user warrant an obvious one. If there were ever to be a system in which the pure application of user warrant was possible, TFR's curated folksonomy would be it. My finding that wranglers considered other forms of evidence is a strong refutation against this possibility. Wranglers valued domain expertise as a classification design skill, often appealing to wrangler discretion to indicate the importance of local, specialized knowledge, but this expertise combined knowledge of user needs along with expertise in source material.

User/Designer Gap

The user/designer gap is a factor of concern in classification theory, with classification scholars taking very different approaches to address the inevitability that designers and users have different views of the classification system. For example, classification designers who advocate for user warrant, such as Bates (1986) and López-Huertas (1997), argue that classification systems should make bridges between user

vocabulary and the vocabulary of the collection. In this perspective, classification systems that incorporate user warrant ameliorate the inevitable gap between users and the collection; classification design is therefore the solution to the gap, and not another source of the gap. The domain analysis approach takes a different approach to the inevitability of this gap between users and the collection by prioritizing literary or consensus warrant; advocates of this approach state that the classification system is an opportunity to teach users the vocabulary of the collection and socialize them into the discourse of the knowledge domain it represents (Hjørland & Albrechtsen, 1995; Jacob, 2001). In this view, there is a gap between users and the classification system, but as users learn to adapt to the classification system, the gap between user terminology and domain terminology closes.

The wranglers' perspectives aligned more closely to the user warrant approach in that wranglers prioritized making the curated folksonomy immediately intelligible to users. The substantial divergence from classification theory is that wranglers continued to consider the possibility that users would not understand the classification system—even when it was constructed from users' own vocabulary. When wranglers reflected that users had a different perspective than wranglers, they noted that users did not have a complete view of each other's decisions. In this sense, the wranglers' perspective aligned more closely with the contemporary criticism of user warrant that users are not a homogenous group and their terminology represents multiple perspectives (Feinberg, 2007; Hjørland, 2013b). As with the operation of hierarchy, the wranglers' experiences of the user/designer gap were embedded in system design choices beyond their control. Specifically, the choice to minimize the intrusiveness of the classification system by unobtrusively redirecting searches for non-synonymous terms to search result pages for canonical terms removed an opportunity to display to users the relationships among tags.

That is, instead of a search for "Superman" leading a user to an intermediate page listing "Superman" as a synonymous term to the canonical tag "Clark Kent," or adding the phrase "redirected from 'Superman'" to the "Clark Kent" result page, the synonymous-to-canonical-tag function was seamless and unobtrusive. This design choice prioritized user experience (by minimizing the users' need to learn the curated folksonomy system) while exacerbating the user/designer gap.

Inclusivity

In classification theory, inclusivity has been a common lens for scholars critiquing existing systems. For example, scholars have focused on topics such as gender and sexuality to reveal the biased and value-laden nature of seemingly "objective" classification systems, demonstrating how these systems fail to be inclusive (Adler, 2012; Berman, 1971; Olson, 2001b). Scholars have also used topics such as marginalized religions and cultures to demonstrate the potential for classification systems to be inclusive of minority perspectives (Olson, 1998; Webster & Doyle, 2008). The current trend in classification theory has been towards more commitment to inclusivity, as scholars turn towards designer agency and responsibility (Feinberg, 2007; Mai, 2010) and explicate the ethics of classification design (Adler & Tennis, 2013; Clare Beghtol, 2002; Fox & Reece, 2012).

Parallel to classification theory, influential infrastructure studies scholars have taken inclusivity as a research lens. Winner's "Do Artifacts Have Politics?", (Winner, 1980) examined the racist rationale and impact of Long Island highway overpasses and revealed the extent to which infrastructure can invisibly instantiate and amplify discriminatory attitudes. Similarly, Star and Bowker's *Sorting Things Out* (Bowker & Star, 1999) revealed the role of race classification in the injustice of South African

apartheid. These works have been particularly influential in infrastructure studies precisely because they connect infrastructure to inclusivity concerns such as race. Recognizing the role that infrastructures such as highways and classification systems play in maintaining—or, potentially, resisting—discrimination has galvanized scholars to consider how these infrastructures are constructed and maintained.

Wranglers appeared to represent perspectives consistent with both these scholarly communities. As in contemporary classification theory, wranglers recognized their responsibility to manage the biases of their classification system. As in infrastructure studies, wranglers indicated awareness of the potential for their classification system to instantiate or resist discrimination. Though these bodies of scholarship may have influenced wranglers' attitudes directly, as some wranglers had academic backgrounds in library and information science, these notions of responsibility, care, and the impact of classification also circulate through the pervasive feminist ethos within the fanwork community. Like the concern for temporality, wranglers' concern for inclusivity represented a belief that imagination and creativity are important skills for classification designers, as designing for inclusivity necessarily demands imagining how the system would appear to users unlike oneself. Furthermore, wranglers' considerations of inclusivity signaled their recognition that justice in the classification system would not come about through consistent application of the rules but required the application of the designers' judgments.

Autocomplete

The majority of information science and design literature on autocomplete has addressed the various mechanisms by which the system might produce more relevant suggestions (Cai & de Rijke, 2016), and only occasionally has examined how users

251

engage with autocomplete suggestions (Mitra, Shokouhi, Radlinski, & Hofmann, 2014). Conversely, wranglers considered the reciprocal effect of autocomplete on user choices, and considered how wranglers' design choices might change the character of the curated folksonomy. The wranglers' sensitivity to the potential impact of autocomplete was more reflective of critical media studies and information studies analysis (Noble & Tynes, 2016) and popular press coverage (Oremus, 2012) of this technical factor than it was of the design field of human-computer interaction. Critical studies and popular press coverage of autocomplete have noted the reciprocal relationship between suggested terms and user beliefs and behavior (Noble & Tynes, 2016) and the role of system designers in gatekeeping language through intentional choices of what appears in autocomplete (Oremus, 2012). Specifically, Oremus (2012) reported on Google designers' delay in removing the term "bisexual" from a list of terms that could not appear in Google search's autocomplete suggestions, despite advocacy groups petitioning for the change. Though wranglers also considered more mundane elements of autocomplete that have been well-represented among the technical literature—such as whether the special characters in "Twenty Øne Piløts" would harmonize with the autocomplete algorithm—their considerations of autocomplete were more likely to overturn decisions based on other factors when they engaged with the critical or popular concern over negative effects of suggesting terms to users.

Server Indexing Burden

The server indexing burden was a factor of classification design that indicated the interrelationship of TFR's layers of infrastructure. One of the primary tenets of infrastructure studies has been the awareness that an infrastructure is always built on an installed base, such as established database structures, and is always embedded within

other technical structures, such that changes in one system will affect another (Star & Ruhleder, 1996). The server indexing burden became a factor in wranglers' decisions when the conceptual work of wrangling—linking tags together—was noticeably entangled in the technical infrastructure of the website's server. Consistent with a third tenet of infrastructure studies—that infrastructures become visible upon breakdown—the relevance of wrangling's underlying infrastructure only became relevant to wranglers when it ceased to work properly.

Like autocomplete, the interaction of the conceptual layer of a classification system (the terms, relationships among terms, and relationships between terms and documents) and the technical layer (the database instantiating these terms and relationships) was a salient factor in wranglers' decisions not predicted in the classification research literature. Infrastructure studies, on the other hand, has predicted the relevance of other technical infrastructures to classification designers' choices.

Summary

This section has demonstrated how the classification design at TFR converges and diverges from the expected information—as warrants—and skills indicated in classification research literature. In addition, I have identified how infrastructure studies has predicted some factors relevant to classification design, including those not emphasized in classification literature. This is not to say that infrastructure studies cannot speak to the remaining factors for which I drew from classification literature for context. Rather, for factors such as hierarchy, classification literature has provided extensive analysis and recommendations appropriate to the field's focus while, for factors such as the server indexing burden, infrastructure studies suggests that the impact of these factors is integral to understanding the system itself. In contrast, classification literature has

253

typically presented factors belonging to the technical context type as peripheral concerns; while the presence of technical factors does not contradict the core findings of classification literature, neither are such factors integrated into our theories of the system itself.

These 10 factors and, to a greater extent, their four types make possible comparisons for competing factors and priorities across classification design contexts. Furthermore, they provide a starting point for analysis across infrastructures and particularly for cyberinfrastructures. The factors within each type are likely to vary according to the kind of infrastructure--particularly those of the first type, definitional factors (ie., the definition of a data repository is different than the definition of a classification system). Similarly, the prioritization of factors by designers will be an important point of comparison across systems and infrastructures. This abstract categorization can facilitate analysis and theory development for infrastructure studies, including but not limited to, classification systems.

This analysis links classification design at TFR to classification and infrastructure design more generally. As a contrasting analysis, focusing on local and situated relevance, I next review how the designers' role at TFR created a system particularly suited to intersecting technological infrastructure and the values of the fanwork community.

DESIGNERS' ROLE IN CLASSIFICATION SYSTEMS

The gap between what the procedures and approaches from classification theory and the character of classification systems is classification design itself. The consequences of classification that critical scholars such as Bowker and Star have identified are not clearly explained by classification theory and the instructional texts this

domain has produced. In this section, I focus on the extent to which classification designers shape the character of classification systems in two ways. First, I look at how an understanding of the technical context of the classification system requires a different view of classification designers' roles and skills. Second, I examine how classification designers create a system that reflects the social world of the collection and its community.

Technical Context Factors in Classification Design

Wranglers' considerations of technical context factors revealed the role of otherwise invisible intersecting infrastructures in determining the character of the classification system. In particular, the role of the factor of the server indexing burden in design decisions indicated the wranglers' sense of the classification system as embedded within the technical infrastructure of the website. That the wranglers compromised on otherwise sound design decisions when those decisions would place substantial burden on the server; wranglers acknowledged that the infrastructure of the classification system was in service to the operation of the site—users posting and reading fanworks—so that a classification system that made the site unusable was worse than none at all. Similar to the server indexing burden, the factor of inclusivity was important to wranglers because they placed the classification system in a broader context, imagining how their decisions might extend and amplify discriminatory language and make potential users feel unwelcome.

The account of wranglers' design decisions surfaces some factors likely to be relevant in many classification systems. The absence of technical context factors in discussions of classification design, in particular, is a glaring example of how classification theory does not predict classification design and, in turn, fails to account for

how classification systems come to be as they are. In particular, leaving out the role of technical context factors in design limits our understanding of the skills designers apply to system design. Whereas in the literature review I contrasted rule-following and creativity as opposite poles in the spectrum of classification design skills, the role of technical context factors suggest that teamwork may be more relevant than is suggested by largely single-person accounts of classification design. Exceptions such as Soergel (1974) and Albrechtsen and Jacob (Albrechtsen & Jacob, 1998) in which the classification designer was among a team of specialists defining the system have suggested a collaboration based on complementary skills. Typically, classification research literature has attended to technological changes by examining new potentials for syntax (López-Huertas, 2008; Mills, 2004; Smiraglia, 2002); the account of design at TFR suggests technological changes will also affect semantic decisions.

The convergence of definitional factors, user-centered factors, and external factors between the TFR wranglers' decisions and classification theory suggests that wranglers' experiences map to classification design in more traditional settings and that technical context factors should also be attended to in classification research. However, we can also learn from looking deeply at the specifics of this situation rather than those elements TFR shares in common with established classification theory. In the following subsection, I examine the social world of TFR to illustrate how classification systems can come to respect and reflect the values of their collections and communities.

The Social World of TFR

If classification systems embody values and bias (Feinberg, 2007; Mai, 2010; Olson, 2001b), it follows that these values have a source in the system's wider context. Classification scholars have argued that relevant contexts include the designers' voices

(Feinberg, 2011a), the designers' philosophical perspectives (Olson, 2001a), and the social world in which designers create the system (Mai, 2010). Here, I focus on Mai's notion of the social world of a classification system and its designers for two reasons. First, the collaborative character of design work at TFR makes an intentional authorial voice unlikely or difficult to identify. Second, and similarly, the lack of a common educational or professional background among the classification designers at TFR suggests that a common philosophical perspective—such as Dewey's Hegelian philosophy (Olson, 2001a)—is less likely to be a determinative force. Instead, the context of the interpretive work designers have done for others—the system's social world—is the most apparent connection between TFR's designers and the values of the classification system.

In traditional classification design settings, it is typical for designers to be information organization professionals embedded in libraries (e.g., Young & Mandelstam, 2013) or providing consulting work for other organizations (e.g., Albrechtsen & Pejtersen, 2003). In the context of TFR, designers were not professionals but interested members of the community who volunteered to maintain its infrastructure. Therefore, more so than in traditional classification design settings, TFR designers were embedded in the social world of the classification system. This embeddedness was not only a function of the wranglers' origin as users; wranglers continued to blur the line between user and designer by appealing to their own perspective as users within debates such as the guideline revision discussions. Contemporary classification design approaches such as participatory design (Albrechtsen & Jacob, 1998) are based on the belief that involving the user community in the design process will create system more consistent with user needs and values (Bjögvinsson, Ehn, & Hillgren, 2012; Carroll, 1996; Sanders & Stappers, 2008). The approach at TFR exemplified this approach by

drawing designers from within the user community. It is therefore not surprising that classification design at TFR clearly reflected the values of its social world.

Among the factors I identified above, the most salient to the fit between classification design and the social world was inclusivity. TFR's position in the social world of the fanwork community explained the wranglers' particular focus on alleviating discrimination and supporting social justice. In other words, wranglers enacted an ethic of care (Fox & Reece, 2012) which "views dilemmas in their specific contexts and accommodates exceptions" (p. 380) to avoid harm. Scholars associate the ethic of care with a feminist perspective (Gilligan, 1982) and contrast it against the justice or universalist ethics (Tronto, 1987) implicit in traditional librarianship and classification design (Fox & Reece, 2012). The group's decision to give wranglers' autonomy over their domains of responsibility and individual wranglers' choices to avoid harm by creating local exceptions distinguishes the TFR approach from universalist classification approaches. TFR was a community-created, non-profit collection in which users contributed to infrastructure (the code, the classification system) and the content (the fanwork themselves). The fanwork community is notably a female-dominated and feminist space (Bury, 2005; Busse, 2015; Fiesler, Morrison, & Bruckman, 2016; Hellekson & Busse, 2006; Katyal, 2006; Stanfill & Condis, 2014). As members of the fanwork community, TFR users were intimately familiar with pervasive, negative perceptions of fanfiction as a marginal, frivolous activity (Jenkins, 2013; Lothian, 2012)—perceptions commonly held about other female-dominated interests (Modleski, 1984). Therefore, it is not surprising that TFR wranglers were committed to avoiding discriminatory decisions that would reproduce these or other biases that harmed their users.

The conflict I observed between user primacy and inclusivity further illustrates the role that the social world of TFR played in shaping classification design and determining the character of the classification system. In addition to being a markedly feminist space, the fanwork community is necessarily a creative space that resists limitations on expression (Fiesler & Bruckman, 2014a; Hadas, 2009; Jenkins, 2013). The value of creative freedom was embodied in the factor of user primacy. Specifically, users created tags with minimal restrictions and wranglers did not alter but only linked these user-created tags; users were free to express themselves through tags without censorship. Instances in which user primacy conflicted with inclusivity in wranglers' decisions represented the uneasy coexistence of the values of creative freedom and the feminist ethic of care. That these instances were not easily resolved but produced contentious debate among the wranglers reflected the state of conflict between these values in TFR's social world. Whether the system should respect and follow user choices, even when those choices could harm others, was a question that engaged wranglers as ethical actors in the design of the system. To create—or follow—strict rules that prioritized user warrant or prescribed ethical warrant would have produced a system that did not reflect the values of the community. Constantly wrestling with how to accommodate both values demanded the skills of creativity and domain expertise rather than rule-following.

The status quo in which wranglers constantly weighed these two values in design decisions was, in itself, consistent with a third value of the fanwork community: the maintenance of the gift economy. Scholars typically identify fanwork communities as gift economies (Hellekson, 2009; Hellekson & Busse, 2006; Stanfill & Condis, 2014). Scholars in this area have used and adapted the anthropological concept of the gift economy—a system of production, sharing, and exchange of valued goods not based on immediate or definite rewards (Mauss, 1990)—to explain the sustainability of members'

voluntary labor (Stanfill & Condis, 2014; Turk, 2014), explore the values of the community (Hellekson, 2009), and account for the role of technology (Pearson, 2007). The most visible and the most recognized gifts have been fanworks—fics (works of fiction), artwork, or vids (fan-edited videos). From a commercial and legal standpoint, it has mattered that fanworks were gifts; the creation and distribution fics and vids as an ongoing exchange of gifts rather than commodities has supported the community's fair use claims against copyright infringement accusations (Fiesler & Bruckman, 2014b; Stanfill & Condis, 2014). However, fics and vids have not been the only gifts or labor exchanged in these spaces. In comparison to the highly circulated and highly visible fanworks, infrastructural work such as classification design is invisible work (Star & Strauss, 1999) that supports the foregrounded exchange of fanworks.

The mode of classification design at TFR—maintenance-oriented, performed by users themselves, and demanding creativity—has been part of the fanwork gift economy. Classification designers could have reduced the constant burden on themselves to wrangle incoming tags and resolve dilemmas had they chose to weigh the factors in their decisions in another way or given their guidelines more authority to determine design decisions. Positioning the main skill of TFR's classification designers as rule-following rather than creativity would have reduced the authority of wrangler discretion and along with it the wranglers' accountability for resolving design dilemmas. Furthermore, TFR could have adopted an automated classification approach (e.g., Golub, Lykke, & Tudhope, 2014; Tsui, Wang, Cheung, & Lau, 2010) that harnessed the corpus of user-created tags or used wrangler decisions as a training set for a machine learning algorithm. TFR also could have adopted the voting approach to curated folksonomy design instantiated at Stack Overflow ("Tag Folksonomy and Tag Synonyms," 2010) and LibraryThing ("Tag combining," n.d.). Either the automated classification approach or

the aggregate voting approach to curated folksonomy design would reduce the burden on classification designers. Instead, classification designers maintained a system in which their contributions were comparable to the creative contributions of fanwork creators (Jenkins, 2013) or the supportive constructive commentary of fanwork readers (Booth, 2015).

These three values of the fanwork community—the feminist ethic of care, creative freedom, and the gift economy—were instantiated in the daily design choices of the wranglers and in their curated folksonomy approach. Classification designers here were making types of decisions common in many classification design settings—deciding on the proper terms for referents and creating relationships among terms—and using information and skills mostly accounted for in classification theory. However, the particular way that these designers instantiated and prioritized these elements of classification design created a system particularly well-suited to the values of the social world of the fanwork community.

Summary

Through this discussion chapter, I have identified the ways in which classification designers' decisions at TFR embodied the information and skills surfaced in classification theory while revealing others for which the literature has not adequately accounted. I also revealed how the classification design decisions were suited to the values of the fanwork community. The visibility of the factors here was partially due to the relative autonomy of classification designers in this setting; absent professional training and in the position to develop and modify their own rules—and even modify the manner in which they modified the rules—wranglers were able to make choices that reflected their understanding of the particular needs of the TFR collection and its

community. Even within the relatively small window of time—three years—in which I observed this community and the two-week periods of time in which wranglers reported their experiences to me, I observed many instances in which wranglers made principled decisions that aligned the character of the classification system to the community's values. It is possible that such remarkable instances occur less often in more strictly controlled classification design settings in which established procedures are a force of inertia, but these instances would still be important to creating and understanding the character of the classification system.

It is also important to note that values are specific to a particular perspective. At TFR, the dual role of wranglers as users and designers meant that the resulting system represented users' values. Within the fanwork context, but from a different perspective, the system may have looked much different. For example, if classification design decisions could only be made by fanwork creators and not readers, factors such as the user/designer gap and authenticity may have been judged differently. An even greater contrast would occur if a fanwork classification system were created from the perspective of source material creators or from designers working for a fanwork website supported by advertising revenue. Some dilemmas might have remained the same (such as name dilemmas around superheroes) while others (such as priority of inclusivity) may have been handled much differently.

Finally, it is important to note that while values and bias are from a particular perspective, designers cannot will their own values into the system. There are mediating factors such as intersecting infrastructures that shape the system and limit the autonomy designers have in determining the character of the classification system. How designers negotiate with these mediating factors and what role classification designers have in pushing back against intersecting infrastructures is also part of the design process. These

should be considered as part of classification design and among the skills the classification designer brings to bear. Combining classification research and infrastructure studies approaches to examining this case made this interplay of designer and technology visible.

In the next chapter, the conclusion, I review the extent to which my research approach surfaced some elements of classification design while backgrounding others. After reviewing these limitations, I suggest several open questions for classification design and information studies that follow from my findings.

Conclusion

In the previous chapter, I illustrated the relationship between classification systems and the values of the community and collection for which designers make the system. Furthermore, I argued that classification designers make judgments that can align the classification system to those local values by diverging from guidelines and by choosing amongst different forms of warrant. I characterized classification designers at the site of my study as reflective and thoughtful, making intentional decisions that prioritized one or more of the established purposes of classification system—to authentically represent an outside reality or to facilitate efficient retrieval—while minimizing harm done. Though I argue that classification designers in any setting are likely to make similarly reflective and intentional decisions, in this chapter I focus on what made classification design at this site unique and draw out four aspects for future research. First, I return to the infrastructure studies lens to draw out what this research lens added to my study of classification design.

THE INFRASTRUCTURE STUDIES RESEARCH LENS

In my findings sections, I illustrated the complexity of interactions among different factors in classification design. Returning to these factors, I discuss the extent to which I made the visibility of these interactions possible by using a research lens combining classification theory and infrastructure studies.

In taking an infrastructure studies approach to understanding classification systems—by focusing on what happens "backstage," by focusing on the minutiae of maintenance work, and by analyzing "breakdowns" in the otherwise seamless operation of the system—I accounted for more of the factors in classification design than I would have by examining the system according to the principles of classification theory. In the

Findings Summary chapter, I noted the prevalence of each factor across the three types of complex design decisions. Here, to illustrate the contributions I have made through the combination of these two research lenses, in Table 6 I repeat the table of 10 factors and indicate to which research tradition each belongs. Eight of the factors I identified are visible through the lens of classification theory; four are visible through the lens of infrastructure studies, including two visible through infrastructure studies alone.

	Classification Theory	Infrastructure Studies
Ambiguity	x	
Filtering	x	
Hierarchy	x	
Temporality	x	x
Authenticity	x	
User primacy	x	
User/designer gap	x	
Inclusivity	x	x
Autocomplete		x
Server indexing burden		x

Table 6 Factors in complex decisions by research lens

While filtering and hierarchy, the first two factors, have been explicit topics of study in classification theory, and the following six have been recurrent themes in the classification theory literature, the last two have not been adequately discussed in

classification theory. My addition of infrastructure studies to the research lens made possible the visibility of two of the factors—the server indexing burden and autocomplete—and their interactions with other factors. Necessarily, the interaction of the server indexing burden with hierarchy, and the interactions between autocomplete and the other factors, would have remained invisible without this combination.

The server indexing burden and autocomplete exemplified the relationship between the conceptual work of classification design—the choice of terms and their relationships—and the technical systems through which these terms are instantiated. The relevance of the server indexing burden to wranglers' decisions revealed the interplay between the classification designer's actions and the functioning of the overall system. The wranglers described the curated folksonomy and their maintenance of it as very important to the function of TFR, but they understood that the curated folksonomy could also interfere with the core functions—posting fanworks and reading fanworks—if it they did not manage their impact on the server.

Conversely, applying the infrastructure studies lens alone would not have revealed the full picture of classification design at TFR. Classification theory was necessary to recognize decisions as decisions—expertise in classification methods facilitated my analysis of design tasks at TFR. Combining the research lenses of classification theory and infrastructure studies facilitated a more complete analysis of classification design than could have been accomplished with either lens alone. Future infrastructure studies research might similarly combine domain expertise into the specific infrastructural system—here a classification system—to extend and deepen analysis. To extend the impact of the infrastructure studies lens, future studies might also situate a given classification system within a network of relationships to other infrastructures. Similar to how the server index and the classification system interacted, TFR's system

also interacted with external classification systems. This interaction was largely indirect, as users carried expectations of classification norms from established websites such as Fanfiction.net and developed new norms as platforms such as tumblr grew in popularity. The above analysis did not capture participants' speculation on the influence of these interactions as managing tumblr-style tags, for example, was a straightforward if time-consuming task for many wranglers. Future analysis that reflects typical rather than complex decisions could capture this factor in classification design. Similarly, data collection that focuses on early negotiations in system design—such as when designers seek to differentiate their system from competitors or facilitate interoperability with peers—would be more likely to reveal the influence of existing, external systems.

Infrastructure studies was a particularly valuable addition to studying classification design as it highlighted the embeddedness of the classification system in other technical infrastructures. In the next two sections, I shift focus to the human side of classification design. First, I review the impact of curated folksonomies as an approach to classification design for user-generated collections. Next, I underline the extent to which curated folksonomy—particularly as TFR instantiated it—represented a turn away from overly computational approaches to classification design.

CURATED FOLKSONOMIES

Curated folksonomies are an interesting, arguably scalable approach to the challenges of large-scale, user-generated collections. They appear to be particularly appropriate knowledge organization approaches for collections that do not have the capital to hire and retain classification design professionals, as the users themselves can contribute small tasks to incrementally improve a folksonomy. Such constant and

incremental design is necessary for collections that are quickly growing and are likely to regularly produce new concepts for integration into the classification system.

The approach of relying on the collections' users to perform design tasks is suited to communities that privilege domain knowledge over professional authority. At TFR, this attitude towards domain knowledge and professional authority was intrinsic to a community that had experienced a history of marginalization in more traditional settings. Users' motivation to contribute to such a system may depend on users' recognition that their domain knowledge is necessary to design a good system, and that professionals are unlikely to share that domain knowledge. Future research on curated folksonomies can investigate the necessary conditions for users to sustain the motivation and effort for the constant maintenance required. Since classification designers' interpretations of the values of their community and the purpose of the classification shape the system it is particularly important to understand why users would engage in this work and who this work attracts. Such research can improve the curated folksonomy approach in other settings and produce social and technical recommendations for future systems. For example, in having users—rather than professionals—design systems, how important is training in classification principles? Specifically, such research could help attune projects to an amount of classification training that is necessary to sustain a good system without dissuading users from engaging in these tasks.

Future research on curated folksonomies can also speak back to classification theory and knowledge organization in general. In particular, the extent that users outside of a professional context converge or diverge from established classification principles is an important area for study. Findings in this area could speak to the tension between classification as a natural activity (e.g., Bliss, 1929; Zerubavel, 1991) and classification as an intentional, specialized activity (C Beghtol, 2003). Similarly, researchers might

consider how the system that the curated folksonomy approach produces differs from one produced through automated classification methods for folksonomies (e.g., Tsui, Wang, Cheung, & Lau, 2010), as these two approaches may become competing solutions to the challenges inherent in organizing large-scale, online collections. In the following section, I focus on the human participation in classification design.

HUMAN CLASSIFICATION DESIGN

Curated folksonomies are a creative solution to the challenges of organizing large and growing collections of user-generated documents. However, they are not the only solution. A more obvious solution—one that is particularly appropriate to large collections and requires far less human labor for its maintenance—would be a machine learning approach that harnesses users' thousands of tagging actions to calculate which tags are equivalent in meaning. Furthermore, such an approach could be built off of the wranglers' work, as their past decisions to designate canonical terms, assign synonymous terms, and create metatag-subtag relationships would provide a rich learning set from which to develop a machine learning solution. Indeed, a growing body of classification research builds attempts to automate classification design using large sets of tags, tag-document relationships, and human judgments of equivalence (Golub et al., 2014).

In light of machine learning approaches to classification system design, the curated folksonomy approach at TFR appeared to devote a great deal of work to avoid a feasible automated approach. This comparison, and the sustained effort of volunteers to maintain the curated folksonomy approach, suggests there is some perceived value in positioning human judgment as the final step of a classification design process

Future research in this area might examine the extent to which users differentiate between algorithmic and human judgment modes of large-scale classification design. In

particular, such research might examine whether systems in which the human judgment is more open and participatory—as in Stack Overflow and LibraryThing's instantiations— are perceived differently than instantiations such as TFR's, in which the decision making is behind-the-scenes. I suggest that, as with other classification design decisions, the extent to which the system makes human judgment open and participatory—or, conversely, hidden and selective—would reflect other values in the community. To that end, in my closing aspect I address the extent to which we might view classification systems and their design as a research lens in their own right.

CLASSIFICATION SYSTEMS AS A RESEARCH LENS

Inherent in my argument that the classification system and the approach to classification design at TFR reflected the values of the community is the idea that the reverse is also true; evidence of the values of TFR's community were found in classification design. Though my questions in this study were about the design process and not about the community, future research could take the reverse approach. Indeed, studies of labelling and naming have provided persuasive and provocative insights into other cultures, such as in Lakoff's (1987) account of gendered nouns in *Women, Fire, and Dangerous Things*. Systematic studies of community values that incorporate studies of how a community agrees to organize and label their knowledge and their documents will reveal a more complete picture of that community.

I argued here that the classification system and its community were complementary, especially as the system reflected the community's value of inclusivity. However, there is the possibility that research into other settings would find less harmonious relationships between system and community. Criticisms of classification systems deployed in contemporary contexts suggest that such findings would be likely in

270

the case of longstanding systems that have been slow to adapt to modern values (e.g., Berman, 1971; Olson, 2001). I suggest that the difference between these two types of findings—systems that are suited to their communities and systems that are not—will be indicated by the extent to which the system is invisible. That is, if users are not aware of the presence of the system in daily use it is because the system matches the users' expectations and needs; therefore, those systems users frequently characterize as frustrating or inconvenient are more likely to embody values contrary to those of the community. Conversely, it may be the case that dominant classification systems are invisible to most users but frustrating to a minority of users. Such cases have the potential to uncover faultlines (Lau & Murnighan, 1998) within communities.

SUMMARY

I have identified four aspects for future research based on the findings of this study. Continuing research on and analysis into classification design has the potential to speak to diverse scholarly communities. Especially as user-generated collections grow and the trend towards end-user rather than mediated interaction with collections continues, insight into how different publics understand and practice classification design will be an important area for information studies in particular.

Appendices

Appendix A: Diary Protocol

ABOUT THIS DIARY:

This diary is part of data collection for a doctoral dissertation on classification design work. It is not endorsed by [TFR], your participation is voluntary, and your answers will not be shared with your fellow wranglers or staff. You may withdraw your participation at any time. You may contact me (Julia Bullard, julia.a.bullard@gmail.com) for clarification at any time.

> The purpose of this diary is to get an idea of what typical wrangling work looks like. Please feel free to answer the questions however briefly (or in whatever form) is most convenient for you and does not overly disrupt your wrangling.

This is an **event-contingent diary protocol** – please answer the diary questions each time you engage in wrangling work (if you have many short wrangling sessions per day, record at least one of them in full and make a note of the others) over the next **two weeks**. The protocol includes a set of questions to help you describe each wrangling session as well as general background and reflection questions that can be answered at any time. The **background** and **reflection** questions only need to be answered once.

Page 2 lists the background questions. Page 3 lists the questions to be completed during wrangling sessions. Page 4 lists the reflection questions to be completed at the end of the two weeks.

BACKGROUND QUESTIONS (ANSWER ONCE):

- How long have you been a wrangler?

- Which fandoms are you responsible for? (if this is a long list, you can give an overview, such as "10 manga fandoms and Teen Wolf (TV)")

- How did you become involved with wrangling? (were you previously an author, a reader, a tag enthusiast?)

This diary is intended to be completed over a **two week** period. Over the course of the two weeks, please create a new, dated entry for each wrangling session (if you have many short wrangling sessions per day, record at least one of them in full and make a note of the others).

Please complete the diary questions as close as possible to the time you do the wrangling work. The questions touch on the start, during, and the end of your session. You may include any form of information – text, screenshots[8], diagrams, links – that helps you answer the questions. To keep the diary from being too much work, or for overly interfering with wrangling, feel free to be as brief in your answers as possible. There will be an opportunity to elaborate and for me to ask for clarification later during a follow-up interview, so enough detail to remember the event would be enough.

For example: simple wrangling task – "Stiles Stilnski" - misspelling of a main character, synned to existing character canonical tag

DIARY QUESTIONS:

- As you begin:
    - What is the date & time?
    - Why now? (Was this session scheduled in advance or did something prompt you to begin work?)
    - What, if anything, did you do in preparation for this session? (Did you do anything to set up your workspace/computer?)
    - What, if anything, is your goal or priority for this session?
- As you wrangle:

---

[8] Walkthroughs on taking screenshots on Mac & Windows: http://www.take-a-screenshot.org/

- o Please keep a record in whatever form is convenient of how you go about your work (for example, record yourself narrating as you go or jot down a description of what you're doing)
- o In particular, please note the following if they occur:
  - An example of an easy or straightforward wrangling task and how you address it
  - An example of a complex wrangling task and how you address it
  - Sources you consult (either within or outside the archive and OTW)
  - Any of your own record-keeping (Excel spreadsheets, notes-to-self)
  - Any communication with other wranglers (campfire, shared spreadsheets, the wiki, the mailing list, private communication)[9]
- After wrangling:
  - o What time did you stop?
  - o Why did you stop now? (did you run out of time, was there a natural stopping point?)
  - o Was this a typical example of your wrangling sessions? (in terms of time, difficulty, methods)

---

[9] No need for names, just that the interaction happened and possibly why

REFLECTION QUESTIONS (ANSWER ONCE, AFTER COMPLETING DIARY):

- Did keeping this diary impact your wrangling? How so?

- Is there anything you usually do while wrangling that this diary didn't cover? That is, is this diary a good representation of your typical wrangling activity or is there something else you usually do but couldn't for some reason (e.g., servers were down, co-wranglers were away)? If so, please describe that activity.

Appendix B: Interview Protocol

Introduction

Thanks for agreeing to this interview. Feel free to let me know if we need to pause or stop the conversation at any time. The questions touch on your understanding of wrangling work and how you go about completing the work. Some of the questions will expand on those from the diary protocol, so we can refer to those as needed. There will be an opportunity at the end to return to any topic you think we've missed. As we talk, if there are any examples you can provide from your own wrangling work, those would really help. You may choose to elaborate on examples from your diary entries or describe others. As a reminder, this interview is part of data collection for a doctoral dissertation on classification design work. It is not endorsed by [the organization], your participation is voluntary, and your answers will not be shared with your fellow wranglers or staff. You may withdraw your participation at any time.

I have some questions about wrangling in general, as well as some about the particular fandoms you have assigned.

Individual Domains

- How do your fandoms differ in terms of wrangling work?
    - Prompt: What types of sources do you consult for different fandoms?
    - Prompt: Are their fandoms you'd describe as busier, or more difficult, or more surprising?
- What's an example of an easy wrangling task in your fandoms?
    - Follow-up: How might you be able to tell that a wrangling task is going to be particularly straightforward?
- What's an example of a difficult wrangling task in your fandoms?
    - Follow-up: How did/would you resolve this issue?
        - Prompt: Are there resources you would check, or experts you would consult?
    - Follow-up: How might you be able to tell that a wrangling task is going to be particularly complex?
    - Follow-up: What would make you feel confident that you had found a good solution to a difficult wrangling task? (Is feeling confident a realistic goal/criterion?)
    - Follow-up: Would you go through the same process or come to the same result in all of your fandoms?
- What would you say are your tools or resources for making difficult wrangling decisions?
    - Follow-up: Have you ever had the experience of these contradicting each other? How do you manage that?

- What does a good syn decision look like? What does a good subtag/metatag decision look like?
  - Prompt: What would make you confident you'd done one of these correctly?
  - Prompt: If a wrangler asked you to check over their work for them, what would you look for?
- [Time permitting, or if lacking specific examples above] What are the particular challenges of _____ fandom? [Take examples from listed fandoms]
  - Prompt: I have a game fandom where the player chooses the gender and name of the main character. That's created a lot of work I don't have in my other domains. Do you have any experiences like that?
- [Time permitting, or if lacking specific examples above] Tell me about the last problem you encountered while wrangling.
  - Follow-up: Why did that happen?
  - Follow-up: How did the problem get resolved?
  - Follow-up: Where did you look for help?

Wrangling Work

- Do you have co-wranglers on any fandoms? If so, how do you communicate with them?
  - Prompt: How do you go about asking for advice or consensus?
  - Prompt: How do you coordinate the workload?
- What would you say is your goal when you tackle unwrangled or unfilterable tags?
  - Prompt: Are there particular queues/bins you prioritize?
  - Prompt: Do you keep track of your progress in any way?
  - Prompt: Is there a type of wrangling task that takes more of your time or attention?
- What do you do when the wrangling interface is shut off?
  - Prompt: [mention most recent outages]
  - Prompt: Are there fannish activities you do during downtime? Are there organizing activities you do during downtime?

Wrangling Project

- How important would you say wrangling is to the functioning of [TFR]?
  - Prompt: How would you describe to a new user the purpose of the wrangling system?
  - Prompt: What would you say are the strengths or weaknesses of the wrangled system compared to pure folksonomy or unregulated tagging (such as Delicious, Tumblr, or Livejournal)?
- Follow-up: Do you think non-wrangler users feel the same way? Why or why not?

- o Prompt: Do you receive any feedback from [TFR] users?
- o Prompt: How do you think the wrangling system affects authors on [TFR]?
- o Prompt: How do you think the wrangling system affects readers on [TFR]?
- [Time permitting] Shortly after I joined, the tag entry pages became visible to all users. Do you think that has had any effect on wrangling? How so / Why not?
  - o Prompt: Has this had any effect on users' tagging behavior?
  - o Prompt: Do you think about how the tag pages will appear to users?

Wrap-Up

- Are there any questions you expected that I haven't asked? Or that you'd ask a wrangler in an interview?
- Is there anything else you'd like to share about your experience wrangling?

Appendix C: Participant Contact Log

Interview Order	Participant Contact	Diary Completed	Interview Conducted
1	2015-06-24	2015-07-11	2015-07-13
2	2015-06-24	2015-07-11	2015-07-15
3	2015-06-27	2015-07-11	2015-07-18
4	2015-06-24	2015-07-17	2015-07-19
5	2015-06-24	2015-07-17	2015-07-22
6	2015-06-24	2015-07-18	2015-08-11
7	2015-06-28	2015-07-18	2015-07-20
8	2015-06-30	2015-07-23	2015-08-06
9	2015-06-24	2015-07-25	2015-07-30
10	2015-06-26	2015-07-28	2015-08-10
11	2015-06-24	2015-08-04	2015-08-26
12	2015-06-25	2015-08-09	2015-08-15
13	2015-07-02	2015-09-29	2015-10-06
14	2015-06-30	2015-11-15	2015-11-22
15	2015-11-02	2015-11-15	2015-11-17
16	2015-11-02	2015-11-18	2015-11-21
17	2015-11-02	2015-11-20	2015-11-23
18	2015-06-28	2015-11-23	2015-11-27
19	2015-11-11	2015-11-26	2015-12-11

References

Adler, M. (2012). *For sexual perversion see paraphilias: Disciplining sexual deviance at the Library of Congress*.

Adler, M., & Tennis, J. T. (2013). Toward a Taxonomy of Harm in Knowledge Organization Systems. *Knowledge Organization*, *40*(4), 266–272.

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

Aitchison, J., Gilchrist, A., & Bawden, D. (2000). *Thesaurus construction and use: A practical manual*. Chicago: Fitzroy Dearborn Publishers.

Albrechtsen, H., & Jacob, E. K. (1998). The dynamics of classification systems as boundary objects for cooperation in the electronic library. *Library Trends*, *47*(2).

Albrechtsen, H., & Pejtersen, A. (2003). Cognitive work analysis and work centered design of classification schemes. *Knowledge Organization*, *30*(3–4), 213–227.

Albrechtsen, H., Pejtersen, A., & Cleal, B. (2002). Empirical work analysis of collaborative film indexing. In *Emerging Frameworks and Methods: Proceedings of the Fourth International Conference on the Conceptions of Library and Information Science*. Seattle, WA.

Bandom (Decaydance+, My Chemical Romance). (n.d.). Retrieved from http://fanlore.org/wiki/Bandom_(Decaydance%2B,_My_Chemical_Romance)

Bandom Terminology Debate. (n.d.). Retrieved from http://fanlore.org/wiki/Bandom_Terminology_Debate

Bates, M. J. (1986). Subject access in online catalogs: A design model. *Journal of the American Society for Information Science*, *37*(6), 357–376. http://doi.org/10.1002/asi.4630370602

Beghtol, C. (1986). Semantic validity: Concepts of warrant in bibliographic classification systems. *Library Resources & Technical Services*, *30*(April/June), 109–125.

Beghtol, C. (1994). *The classification of fiction: The development of a system based on theoretical principles*. Metuchen, New Jersey: Scarecrow Press.

Beghtol, C. (1995). Domain analysis, literary warrant, and consensus: the case of fiction studies. *Journal of the American Society for Information Science*, *46*(1), 30–44.

Beghtol, C. (2002). A proposed ethical warrant for global knowledge representation and organization systems. *Journal of Documentation*, *58*(5), 507–532. http://doi.org/10.1108/00220410210441

Beghtol, C. (2003). Classification for information retrieval and classification for knowledge discovery: Relationships between "professional" and "naive" classification. *Knowledge Organization*, *30*(2), 64–73.

Bennett, R., Lavoie, B. F., & O'Neill, E. T. (2003). The concept of a work in WorldCat: An application of FRBR. *Library Collections, Acquisition and Technical Services*, *27*(1), 45–59. http://doi.org/10.1016/S1464-9055(02)00306-8

Berman, S. (1971). *Prejudices and antipathies: A tract on the LC subject heads concerning people*. Metuchen, New Jersey: Scarecrow Press.

Bernard, H. R. (2006). *Research Methods in Anthropology: Qualitative and Quantitative Approaches* (Vol. 4th). http://doi.org/10.1525/aa.2000.102.1.183

Bioinformatics & Research Computing at Whitehead Institute. (2009). Venn Diagram Generator. Retrieved from http://jura.wi.mit.edu/bioc/tools/venn.php

Bjögvinsson, E., Ehn, P., & Hillgren, P.-A. (2012). Design Things and Design Thinking: Contemporary Participatory Design Challenges. *Design Issues*, *28*(3), 101–116. http://doi.org/10.1162/DESI_a_00165

Bliss, H. E. (1929). *The Organization of Knowledge and the System of the Sciences*. New York, NY: Holt & Co.

Booth, P. J. (2015). Fandom: The classroom of the future. *Transformative Works and Cultures*, *19*.

Bowker, G. C., Baker, K., Millerand, F., & Ribes, D. (2010). Toward information infrastructure studies: Ways of knowing in a networked environment. In J. Hunsinger, L. Klastrup, & M. Allen (Eds.), *International Handbook of Internet Research* (pp. 97–117). Dordrecht: Springer Netherlands. http://doi.org/10.1007/978-1-4020-9789-8

Bowker, G. C., & Star, S. L. (1999). *Sorting Things Out: Classification and Its Consequences*. Cambridge: MIT Press.

Broughton, V. (2006). The need for a faceted classification as the basis of all methods of information retrieval. *Aslib Proceedings*, *58*(1/2), 49–72. http://doi.org/10.1108/00012530610648671

Broughton, V., & Slavic, A. (2007). Building a faceted classification for the humanities: principles and procedures. *Journal of Documentation*, *63*(5), 727–754. http://doi.org/10.1108/00220410710827772

Buchanan, R. (1992). Wicked problems in design thinking. *Design Issues*, *8*(2), 5–21.

Buckland, M. (1999). Vocabulary as a central concept in library and information science. In T. Arpanac (Ed.), *Digital Libraries: Interdisciplinary Concepts, Challenges, and Opportunities. Proceedings of the Third International Conference on Conceptions of Library and Information Science* (pp. 3–12). Dubrovnik, Croatia.

Buckland, M. (2012). Obsolescence in Subject Description. *Journal of Documentation*, *68*.

Bullard, J. (2016). Motivating invisible contributions: Framing volunteer classification

design in a fanfiction repository. In *Proceedings of the 19th International Conference on Supporting Group Work - GROUP '16* (pp. 181–193). New York, New York, USA: ACM Press. http://doi.org/10.1145/2957276.2957295

Bury, R. (2005). *Cyberspaces of Their Own: Female Fandoms Online*. New York: Peter Lang.

Busse, K. (2015). Fan Labor and Feminism: Capitalizing on the Fannish Labor of Love. *Cinema Journal*, *54*(3), 110–115. http://doi.org/10.1353/cj.2015.0034

Cai, F., & de Rijke, M. (2016). A Survey of Query Auto Completion in Information Retrieval. *Foundations and Trends® in Information Retrieval*, *10*(4), 273–363. http://doi.org/10.1561/1500000055

Carroll, J. M. (1996). Encountering others: Reciprocal openings in participatory design and user-centered design. *Human–Computer Interaction*, *11*(3), 285–290.

Carter, S., & Mankoff, J. (2005). When participants do the capturing: The role of media in diary studies. In *Proceedings of the SIGCHI conference on Human Computer Interaction* (pp. 899–908).

Clark, A. (1998). *Being there: Putting brain, body, and world together again*.

Cleverdon, C. W., & Keen, M. (1966). Factors determining the performance of indexing systems. *Aslib Cranfield Research Project*, *2. Test Re*.

Coogan, P. (2009). The Definition of the Superhero. *A Comics Studies Reader*.

Creswell, J., & Miller, D. (2000). Determining validity in qualitative inquiry. *Theory into Practice*, *39*(3), 124–130. http://doi.org/10.1207/s15430421tip3903

Cutter, C. A. (1891). *Expansive classification*. Boston.

Cutting, D. R., Karger, D. R., Pedersen, J. O., & Tukey, J. W. (1992). Scatter/Gather: a cluster-based approach to browsing large document collections. In *Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '92* (pp. 318–329). New York, New York, USA: ACM Press. http://doi.org/10.1145/133160.133214

Dewey, M. (1876). *A classification and subject index for cataloguing and arranging the books and pamphlets of a library*. Hartford, Conn.: Case, Lockwood & Brainard Company.

Edwards, P. N. (2004). Infrastructure and modernity: Force, time, and social organization in the history of sociotechnical systems. *Modernity and Technology*, 185–225. http://doi.org/0262633108

Edwards, P. N., Jackson, S. J., Bowker, G. C., & Knobel, C. (2007). Understanding infrastructure: Dynamics, tensions, and design. Report of a Workshop on "History & Theory of Infrastructure: Lessons for New Scientific Cyberinfrastructures." *Design*, (January), 54. http://doi.org/2027.42/49353

Ellison, H. (2012, August 13). The book burning that wasn't: Thousands of works of fiction destroyed and no one pays attention. *The Huffington Post*.

Ereshefsky, M. (2000). *The Poverty of the Linnaean Hierarchy: A Philosophical Study of Biological Taxonomy*. Cambridge: Cambridge University Press.

Feinberg, M. (2007). Hidden bias to responsible bias: an approach to information systems based on Haraway's situated knowledges. *Information Research*, *12*(4).

Feinberg, M. (2011a). How information systems communicate as documents: the concept of authorial voice. *Journal of Documentation*, *67*(6), 1015–1037. http://doi.org/10.1108/00220411111183573

Feinberg, M. (2011b). Organization as expression: Classification as digital media. In B. Aspray & M. Winget (Eds.), *Digital Media*. Scarecrow Press.

Feinberg, M., Carter, D., & Bullard, J. (2014a). A Story Without End : Writing the Residual into Descriptive Infrastructure. In *DIS '14 Proceedings of the Designing Interactive Systems Conference* (pp. 385–394). http://doi.org/10.1145/2598510.2598553

Feinberg, M., Carter, D., & Bullard, J. (2014b). Always somewhere, never there: Using critical design to understand database interactions. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14* (pp. 1941–1950). New York, New York, USA: ACM Press. http://doi.org/10.1145/2556288.2557055

Fidel, R. (1994). User-centered indexing. *Journal of the American Society for Information Science*, *45*(8), 572--576.

Fidel, R., & Pejtersen, A. (2004). From information behaviour research to the design of information systems: the Cognitive Work Analysis framework. *Information Research*, *10*(1).

Fiesler, C., & Bruckman, A. (2014a). Copyright terms in online creative communities. In *Proceedings of CHI 2014* (pp. 2551–2556). http://doi.org/10.1145/2559206.2581294

Fiesler, C., & Bruckman, A. S. (2014b). Remixers' understandings of fair use online. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing - CSCW '14* (pp. 1023–1032). New York, New York, USA: ACM Press. http://doi.org/10.1145/2531602.2531695

Fiesler, C., Morrison, S., & Bruckman, A. S. (2016). An Archive of Their Own: A Case Study of Feminist HCI and Values in Design. *CHI '16 Proceedings of the ACM Conference on Human Factors in Computing Systems*. http://doi.org/10.1145/2858036.2858409

Fleischmann, K. R., & Wallace, W. a. (2007). Ethical implications of values embedded in computational models: An exploratory study. *Proceedings of the American Society for Information Science and Technology*, *43*(1), 1–16.

http://doi.org/10.1002/meet.14504301254

Foskett, D. J. (1974). *Classification and indexing in the social sciences* (2nd ed.). London: Butterworths.

Fox, M., & Reece, A. (2012). Which ethics? Whose morality?: An analysis of ethical standards for information organization. *Knowledge Organization*, *39*(5), 377–383.

Friedman, A., & Thellefsen, M. (2011). Concept theory and semiotics in knowledge organization. *Journal of Documentation*, *67*(4), 644–674. http://doi.org/10.1108/00220411111145034

Furner, J. (2004). Information studies without information. *Library Trends*, *52*(3), 427–446.

Galison, P. (1996). Computer simulations and the trading zone. *The Disunity of Science: Boundaries, Contexts, and Power*, 118–157.

Gergen, K. (1992). *The saturated self: Dilemmas of identity in contemporary life*.

Gilchrist, A. (2003). Thesauri, taxonomies and ontologies – an etymological note. *Journal of Documentation*, *59*(1), 7–18. http://doi.org/10.1108/00220410310457984

Gilligan, C. (1982). In a Different Voice: Psychological Theory and Women's Development. Cambridge, Mass.: Harvard University Press.

Glaser, B. G., & Strauss, A. L. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Hawthorne, NY: Aldine de Gruyter.

Golder, S. a., & Huberman, B. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science*, *32*(2), 198–208. http://doi.org/10.1177/0165551506062337

Golder, S., & Huberman, B. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science*, *32*(August 2005), 198–208. http://doi.org/10.1177/0165551506062337

Golub, K. (2006). Automated subject classification of textual web documents. *Journal of Documentation*, *62*(3), 350–371. http://doi.org/10.1108/00220410610666501

Golub, K., Lykke, M., & Tudhope, D. (2014). Enhancing social tagging with automated keywords from the Dewey Decimal Classification. *Journal of Documentation*, *70*(5), 801–828. http://doi.org/10.1108/JD-05-2013-0056

Gould, J. D., & Lewis, C. (1985). Designing for usability: key principles and what designers think. *Communications of the ACM*, *28*(3), 300–311.

Gratton, C. (2016). Resisting the Gender Binary: The Use of (ING) in the Construction of Non-binary Transgender Identities. *University of Pennsylvania Working Papers in Linguistics*, *22*(2).

Hadas, L. (2009). The web planet: How the changing Internet divided Doctor Who fan

fiction writers. *Transformative Works and Cultures*, *3*(3). http://doi.org/10.3983/twc.2009.0129

Haimson, O. L., Brubaker, J. R., Dombrowski, L., & Hayes, G. R. (2016). Digital Footprints and Changing Networks During Online Identity Transitions. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16* (pp. 2895–2907). New York, New York, USA: ACM Press. http://doi.org/10.1145/2858036.2858136

Hellekson, K. (2009). A Fannish Field of Value: Online Fan Gift Culture. *Cinema Journal*, *48*(4), 113–118. http://doi.org/10.1353/cj.0.0140

Hellekson, K., & Busse, K. (2006). *Fan fiction and fan communities in the age of the internet*. McFarland.

Hidderley, R., & Rafferty, P. (1997). Democratic Indexing: An approach to the retrieval of Fiction. *Information Services & Use*, *17*(2–3), 101–109. http://doi.org/10.3233/ISU-1997-172-304

Hjørland, B. (1992). The concept of "subject" in information science. *Journal of Documentation*, *48*(2), 172–200.

Hjørland, B. (1998). The classification of psychology: A case study in the classification of a knowledge field. *Knowledge Organization*, *25*(4), 162–201.

Hjørland, B. (2002). Domain analysis in information science: Eleven approaches – traditional as well as innovative. *Journal of Documentation*, *58*(4), 422–462. http://doi.org/10.1108/00220410210431136

Hjørland, B. (2004). Arguments for philosophical realism in library and information science. *Library Trends*, *52*(3), 488–506.

Hjørland, B. (2013a). Theories of knowledge organization—theories of knowledge. *Knowledge Organization*, *40*(3), 169–181.

Hjørland, B. (2013b). User based and cognitive approaches to knowledge organization. *Knowledge Organization*, *40*(1), 11–27.

Hjørland, B., & Albrechtsen, H. (1995). Toward a New Horizon in Information Science: Domain-Analysis. *Journal of the American Society for Information Science*, *46*(6), 400–425.

Hoffman, G. (2009). Applying the User-Centered Paradigm to Cataloging Standards in Theory and Practice: Problems and Prospects. *NASKO*, *2*, 27–34.

Hulme, E. W. (1911). Principles of book classification, Part 3. *Library Association Record*, *13*, 444–449.

Hutchins, W. J. (1978). The concept of "aboutness" in subject indexing. *From: Aslib Proceedings*, *30*(5), 172–181. http://doi.org/10.1108/eb050629

Huvila, I. (2006). *The Ecology of Information Work: A Case Study of Bridging*

Archaeological Work and Virtual Reality Based Knowledge Organization. Abo Akademi.

Ibekwe-SanJuan, F. (2006). Constructing and maintaining knowledge organization tools: a symbolic approach. *Journal of Documentation*, *62*(2), 229–250. http://doi.org/10.1108/00220410610653316

Iida, M., Shrout, P., Laurenceau, J., & Bolger, N. (2012). Using Diary Methods in Psychological Research. *APA Handbook of Research Methods in Psychology: Vol. 1. Foundations, Planning, Measures and Psychometrics*. http://doi.org/10.1037/13619-016

Jacob, E. K. (2001). The everyday world of work: Two approaches to the investigation of classification in context. *Journal of Documentation*, *57*(1), 76–99.

Jacob, E. K. (2004). Classification and categorization: A difference that makes a difference. *Library Trends*, *52*(3), 515–540.

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM Computing Surveys*, *31*(3), 264–323. http://doi.org/10.1145/331499.331504

Jenkins, H. (2013). *Textual poachers: Television fans and participatory culture* (20th Anniv). New York: Routledge.

Joint Steering Committee for Revision of AACR., American Library Association., Canadian Library Association., & Chartered Institute of Library and Information Professionals (Great Britain). (2003). *Anglo-American cataloguing rules* (2nd Editio). Canadian Library Association.

Katyal, S. K. (2006). Performance, property, and the slashing of gender in fan fiction. *The American University Journal of Gender, Social Policy & the Law*, *14*(3), 461–518. http://doi.org/10.1525/sp.2007.54.1.23.

Keshet, Y. (2011). Classification systems in the light of sociology of knowledge. *Journal of Documentation*, *67*(1), 144–158. http://doi.org/10.1108/00220411111105489

Knorr-Cetina, K. (1999). *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge, Mass.: Harvard University Press.

Kublik, A., Clevette, V., Ward, D., & Olson, H. A. (2003). Adapting dominant classifications to particular contexts. *Cataloging & Classification Quarterly*, *37*(1–2), 13–31. http://doi.org/10.1300/J104v37n01

Kuhn, T. S. (1964). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Labrou, Y., & Finin, T. (1999). Yahoo! as an ontology. In *Proceedings of the eighth international conference on Information and knowledge management - CIKM '99* (pp. 180–187). New York, New York, USA: ACM Press. http://doi.org/10.1145/319950.319976

Lakoff, G. (UC B. (1987). *Women, Fire, and Dangerous Things*. Chicago: University of Chicago Press.

Langridge, D. W. (1976). *Classification and indexing in the humanities*. London: Butterworths.

Lau, D. C., & Murnighan, J. K. (1998). Demographic diversity and faultlines: The compositional dynamics of organizational groups. *Academy of Management Review*, *23*(2), 325–340. http://doi.org/10.5465/AMR.1998.533229

Lee, C. P., Dourish, P., & Mark, G. (2006). The human infrastructure of cyberinfrastructure. *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work - CSCW '06*, 483. http://doi.org/10.1145/1180875.1180950

Levinson, J. (1980). What a musical work is. *Journal of Philosophy*, *77*(1), 5–28. http://doi.org/10.2307/2025596

Lipscomb, C. E. (2000). Medical Subject Headings (MeSH). *Bulletin of the Medical Library Association*, *88*(3), 265–6.

López-Huertas, M. J. (1997). Thesaurus structure design: a conceptual approach for improved interaction. *Journal of Documentation*, *53*(2), 139–177.

López-Huertas, M. J. (2008). Some current research questions in the field of knowledge organization. *Knowledge Organization*, *35*(2), 113–136.

Lothian, A. (2012). Archival anarchies: Online fandom, subcultural conservation, and the transformative work of digital ephemera. *International Journal of Cultural Studies*, *16*(6), 541–556. http://doi.org/10.1177/1367877912459132

Mai, J.-E. (2005). Analysis in indexing: document and domain centered approaches. *Information Processing & Management*, *41*(3), 599–611. http://doi.org/10.1016/j.ipm.2003.12.004

Mai, J.-E. (2010). Classification in a social world: bias and trust. *Journal of Documentation*, *66*(5), 627–642. http://doi.org/10.1108/00220411011066763

Markey, K. (2006). Forty years of classification online: Final chapter or future unlimited? *Cataloging & Classification Quarterly*, *42*(3–4), 1–63. http://doi.org/10.1300/J104v42n03

Mauss, M. (1990). *The Gift: The Form and Reason for Exchange in Archaic Societies*. New York: W.W. Norton.

Mayor, C. (2012). *The classification of gene products in the molecular biology domain : realism , objectivity , and the limitations of the Gene Ontology*. City University London.

McCallum, A. K., Nigam, K., Rennie, J., & Seymore, K. (n.d.). Automating the Construction of Internet Portals with Machine Learning. *Information Retrieval*, *3*(2),

127–163. http://doi.org/10.1023/A:1009953814988

McKenzie, P. J. (2003). A model of information practices in accounts of everyday-life information seeking. *Journal of Documentation*, *59*(1), 19–40. http://doi.org/10.1108/00220410310457993

Millerand, F., & Bowker, G. C. (2007). Metadata standards: Trajectories and Enactment in the Life of an Ontology. In S. L. Star & M. Lampland (Eds.), *Formalizing Practices: Reckoning with Standards, Numbers and Models in Science and Everyday Life*. http://doi.org/Article

Mills, J. (1964). *The Universal Decimal Classification*. New Brunswick, New Jersey: Rutgers University Press.

Mills, J. (2004). Faceted classification and logical division in information retrieval. *Library Trends*, *52*(3), 541–570.

Mitra, B., Shokouhi, M., Radlinski, F., & Hofmann, K. (2014). On user interactions with query auto-completion. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval - SIGIR '14* (pp. 1055–1058). New York, New York, USA: ACM Press. http://doi.org/10.1145/2600428.2609508

Modleski, T. (1984). *Loving with a Vengeance: Mass-Produced Fantasies for Women*. New York: Methuen.

Morse, J. M. (2004). Theoretical Saturation. In *Encyclopedia of Social Science Research Methods*. 2455 Teller Road, Thousand Oaks California 91320 United States: {SAGE} Publications, Inc.

Munk, T. B., & Mørk, K. (2007). Folksonomy, the power law & the significance of the least effort. *Knowledge Organization*, *34*(1), 16–33.

National Information Standards Organization. (2005). *ANSI/NISO Z39.19-2005 (R2010): Guidelines for the construction, format, and management of monolingual controlled vocabularies* (Revised 20). NISO Press.

Nero, L. (2006). Classifying the popular music of Trinidad and Tobago. *Cataloging & Classification Quarterly*, *42*(3–4), 119–133. http://doi.org/10.1300/J104v42n03

Nielsen, M. (2001). A framework for work task based thesaurus design. *Journal of Documentation*, *57*(6), 774–797.

Noble, S. U., & Tynes, B. M. (2016). *The intersectional Internet : race, sex, class and culture online*. Peter Lang Publishing, Inc.

Norman, D. A., & Draper, S. W. (1986). User centered system design. *New Perspectives on Human-Computer Interaction, L. Erlbaum Associates Inc., Hillsdale, NJ*.

O'Neill, E. T. (2002). FRBR: Functional Requirements for Bibliographic Records. Application of the entity-relationship model to Humphry Clinker. *Library Resources*

and Technical Services, *46*(4), 150–159.
http://doi.org/http://dx.doi.org/10.5860/lrts.46n4.150

Ohly, S., Sonnentag, S., Niessen, C., & Zapf, D. (2010). Diary Studies in Organizational Research: An Introduction and Some Practical Recommendations. *Journal of Personnel Psychology*, *9*(2), 79–93. http://doi.org/10.1027/1866-5888/a000009

Olson, H. A. (1998). Mapping beyond Dewey's boundaries: Constructing classificatory space for marginalized knowledge domains. *Library Trends*, *47*(2), 3–20.

Olson, H. A. (2000). Difference, culture and change: The untapped potential of LCSH. *Cataloging & Classification Quarterly*, *29*(1–2), 53–71.

Olson, H. A. (2001a). Sameness and Difference. *Library Resources & Technical Services*, *45*(3), 115–123.

Olson, H. A. (2001b). The Power to Name: Representation in Library Catalogs. *Signs*, *26*(3), 639–668.

Olson, H. A. (2004). The ubiquitous hierarchy: An army to overcome the threat of a mob. *Library Trends*, *52*(3), 604–616.

Olson, H. A., & Schlegl, R. (2001). Standardization, objectivity, and user focus: A meta-analysis of subject access critiques. *Cataloging & Classification Quarterly*, (February 2012), 37–41.

Oremus, W. (2012, September 13). Google blacklist: Bisexual banned from autocomplete no longer. *Slate Magazine*.

Otlet, P. (1990). *International organisation and dissemination of knowledge: Selected essays of Paul Otlet.* (W. B. Rayward, Ed.). New York: Elsevier.

Park, O. N. (2008a). *Current practice in classification system design: An empirical investigation of classification system design team practice*. University of Washington.

Park, O. N. (2008b). Opening ontology design: A study of the implications of knowledge organization for ontology design. *Knowledge Organization*, *35*(4), 209–221.

Pearson, E. (2007). Digital gifts: Participation and gift exchange in Livejournal communities. *First Monday*, *12*(5).

Peponakis, M. (2012). Conceptualizations of the Cataloging Object: A Critique on Current Perceptions of FRBR Group 1 Entities. *Cataloging & Classification Quarterly*, *50*(5–7), 587–602. http://doi.org/10.1080/01639374.2012.681275

Peters, I., & Weller, K. (2008). Tag gardening for folksonomy enrichment and maintenance. *Webology*, *5*(3).

Rafferty, P., & Hidderley, R. (2007). Flickr and Democratic Indexing: dialogic approaches to indexing. *ASLIB Proceedings*, *59*(4), 397–410. http://doi.org/10.1108/00012530710817591

Ranganathan, S. R. (1959). *Elements of Library Classification*. (B. I. Palmer, Ed.). London: The Association of Assistant Librarians.

Ranganathan, S. R. (1961). *Reference Service*. Bombay: Asia Publishing House.

Ranganathan, S. R. (1962). *Elements of Library Classification*. New York: Asia Publishing House.

Razikin, K., Goh, D. H., Chua, A. Y. K., & Lee, C. S. (2011). Social tags for resource discovery: a comparison between machine learning and user-centric approaches. *Journal of Information Science*, *37*(4), 391–404. http://doi.org/10.1177/0165551511408847

Renear, A., & Dubin, D. (2003). Towards Identity Conditions for Digital Documents. *International Conference on Dublin Core and Metadata Applications*, 181–189.

Ribes, D. (2014). Ethnography of scaling, or, how to a fit a national research infrastructure in the room. *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '14*, 158–170. http://doi.org/10.1145/2531602.2531624

Ribes, D., & Bowker, G. C. (2009). Between meaning and machine: Learning to represent the knowledge of communities. *Information and Organization*, *19*, 199–217. http://doi.org/10.1016/j.infoandorg.2009.04.001

Ribes, D., & Lee, C. P. (2010). Sociotechnical studies of cyberinfrastructure and e-research: Current themes and future trajectories. *Computer Supported Cooperative Work*, *19*, 231–244. http://doi.org/10.1007/s10606-010-9120-0

Rieh, S. Y. (2005). Cognitive Authority. In K. E. Fisher, S. Erdelez, & L. McKechnie (Eds.), *Theories of Information Behavior*. Melford, N.J.: American Society for Information Science and Technology.

Rosch, E., & Mervis, C. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *605*, 573–605.

Rubin, H., & Rubin, I. (2005). *Qualitative interviewing: the art of hearing data.* (2nd Editio). Thousand Oaks, CA: Sage Publications.

Sanders, E. B.-N., & Stappers, P. J. (2008). Co-creation and the new landscapes of design. *Co-Design*, *4*(1), 5–18.

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys*, *34*(1), 1–47. http://doi.org/10.1145/505282.505283

Shearer, J. (2004). A practical exercise in building a thesaurus. *Cataloging & Classification Quarterly*, (August 2013), 37–41. http://doi.org/10.1300/J104v37n03

Smiraglia, R. P. (2001). *The nature of "a work" : implications for the organization of knowledge*. Lanham, Md.: Scarecrow Press.

Smiraglia, R. P. (2002). The progress of theory in knowledge organization. *Library*

Trends, *50*(3), 330–349.

Smith, G. (2009). The superhero as labor. *The Contemporary Comic Book Superhero*.

Soergel, D. (1974). *Indexing languages and thesauri: Construction and maintenance*. Los Angeles: Melville Publishing Company.

Some LibraryThing Concepts. (n.d.). Retrieved February 8, 2016, from http://www.librarything.com/concepts#what

Spärck Jones, K. (1970). Some thoughts on classification for retrieval. *Journal of Documentation*, *26*(2), 89–101. http://doi.org/10.1108/eb026488

Stanfill, M., & Condis, M. (2014). Fandom and/as labor. *Transformative Works and Cultures*, *15*.

Star, S. L. (1999). The Ethnography of Infrastructure. *American Behavioral Scientist*, *43*(3), 377–391. http://doi.org/10.1177/00027649921955326

Star, S. L., & Ruhleder, K. (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information Systems Research*, *7*(1), 111–134.

Star, S. L., & Strauss, A. (1999). Layers of silence, arenas of voice: The ecology of visible and invisible work. *Computer Supported Cooperative Work (CSCW)*, (1995), 9–30.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* .

Svenonius, E. (2000). LCSH: Semantics, syntax and specificity. *Cataloging & Classification Quarterly*, *29*(1–2), 17–30.

Szostak, R. (2008a). Classification, interdisciplinarity, and the study of science. *Journal of Documentation*, *64*(3), 319–332. http://doi.org/10.1108/00220410810867551

Szostak, R. (2008b). Classification, interdisciplinarity, and the study of science. *Journal of Documentation*, *64*(3), 319–332. http://doi.org/10.1108/00220410810867551

Tag combining. (n.d.). Retrieved February 8, 2016, from http://www.librarything.com/wiki/index.php/Tag_combining

Tag Folksonomy and Tag Synonyms. (2010). Retrieved February 8, 2016, from https://blog.stackoverflow.com/2010/08/tag-folksonomy-and-tag-synonyms/

Theimer, S. (2012). A cataloger's resolution to become more creative: How and why. *Cataloging & Classification Quarterly*, *50*(8), 894–902. http://doi.org/10.1080/01639374.2012.711440

Thornton, K., & McDonald, D. W. (2012). Tagging Wikipedia: Collaboratively creating a category system. In *Proceedings of the 2012 ACM Conference on Supporting Group Work* (pp. 219–227).

Tillett, B. B. (2005). FRBR and Cataloging for the Future. *Cataloging & Classification*

Quarterly, *39*(3–4), 197–205. http://doi.org/10.1300/J104v39n03_12

Trimmer, J. D. (1980). The present situation in quantum mechanics: A translation of Schrödinger's "Cat Paradox." *Proceedings of the American Philosophical Society*, *124*(5), 323–338.

Tronto, J. C. (1987). Beyond Gender Difference to a Theory of Care. *Source: Signs*, *12*(4), 644–663.

Tsui, E., Wang, W. M., Cheung, C. F., & Lau, A. S. M. (2010). A concept-relationship acquisition and inference approach for hierarchical taxonomy construction from tags. *Information Processing and Management*, *46*(1), 44–57. http://doi.org/10.1016/j.ipm.2009.05.009

Tudhope, D., & Binding, C. (2008). Faceted Thesauri. *Axiomathes*, *18*(2), 211–222. http://doi.org/10.1007/s10516-008-9031-6

Turk, T. (2014). Fan work: Labor, worth, and participation in fandom's gift economy. *Transformative Works and Cultures*, *15*. http://doi.org/10.3983/twc.2014.0518

van Rijsbergen, C. J. (1979). *Information Retrieval* (2nd ed.). London: Butterworths.

Vickery, B. C. (1960). *Faceted classification: A guide to construction and use of special schemes*. London: Aslib.

Vickery, B. C. (1964). *Classification and indexing in science*. London: Butterworths.

Wacholder, N., Evans, D. K., & Klavans, J. L. (2001). Automatic identification and organization of index terms for interactive browsing. In *Proceedings of the first ACM/IEEE-CS joint conference on Digital libraries - JCDL '01* (pp. 126–134). New York, New York, USA: ACM Press. http://doi.org/10.1145/379437.379468

Wang, Z., Chaudhry, A. S., & Khoo, C. S. G. (2008). Using classification schemes and thesauri to build an organizational taxonomy for organizing content and aiding navigation. *Journal of Documentation*, *64*(6), 842–876. http://doi.org/10.1108/00220410810912424

Webster, K., & Doyle, A. (2008). Don't class me in antiquities! Giving voice to Native American materials. In K. R. Roberto (Ed.), *Radical Cataloging: Essays at the Front*. Jefferson, N.C.: McFarland.

Wild, P. J., Giess, M. D., & McMahon, C. a. (2009). Describing engineering documents with faceted approaches: Observations and reflections. *Journal of Documentation*, *65*(3), 420–445. http://doi.org/10.1108/00220410910952410

Wilson, P. (1983). *Second-hand knowledge: An inquiry into cognitive authority*. Westport, Conn.: Greenwood Press.

Winner, L. (1980). Do artifacts have politics? *Daedalus*, *109*(1), 121–136.

Wittgenstein, L. (1958). *Philosophical investigations*. Oxford: Blackwell.

Yi, K., & Mai Chan, L. (2009). Linking folksonomy to Library of Congress subject headings: an exploratory study. *Journal of Documentation*, *65*(6), 872–900. http://doi.org/10.1108/00220410910998906

Young, J. L., & Mandelstam, Y. (2013). It takes a village: Developing library of congress genre/form terms. *Cataloging & Classification Quarterly*, *51*(1–3), 6–24. http://doi.org/10.1080/01639374.2012.715117

Zeng, M. L. (2008). Knowledge organization Systems. *Knoweldge Organization*, *35*(2–3), 160–182. http://doi.org/10.1007/978-0-387-39940-9_2929

Zerubavel, E. (1991). *The Fine Line: Making Distinctions in Everyday Life*. New York: The Free Press.

Zhang, Y., Zhang, J., Lease, M., & Gwizdka, J. (2014). Multidimensional relevance modeling via psychometrics and crowdsourcing. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval - SIGIR '14* (pp. 435–444). New York, New York, USA: ACM Press. http://doi.org/10.1145/2600428.2609577

Zimmerman, D., & Wieder, D. (1977). The Diary:" Diary-Interview Method" *Journal of Contemporary*.