ISSUE NO. 8

Women's Ways of Structuring Data

Christine L. Masters

Smoothly functioning infrastructures are invisible. Examples of infrastructures range from those physically constructed, such as transportation and public utility systems, to those that are more elusive or fluctuating—systems of economic exchange, for example. When systems work well, people do not realize their immersion within them because they facilitate the ease of daily experiences. For example, we are not always aware of how much we rely on the power grid until a transformer breakdown causes our lights to go out. Infrastructures are complex and sometimes require work to understand and map out, yet once we are aware of how they exist, we find it hard to believe how we could have overlooked them in the first place. According to Geoffrey Bowker and Susan Leigh Star (1999), "the trick [to seeing infrastructure] is to question every apparently natural easiness in the world around us and look for the work involved in making it easy" (p. 39). A definition of infrastructure has several qualities: "embeddedness," "transparency," "reach or scope," "learned as part of membership," "links with conventions of practice," "embodiment of standards," "built on an installed base," "becomes visible upon breakdown," and "is fixed in modular increments, not all at once or globally" (Star & Ruhleder, 1996). Information systems scholars have examined infrastructures within a variety of contexts, working towards revealing both their material and their symbolic natures.

Just as infrastructures themselves are often invisible, women's roles within them have been rendered even more invisible. Whether or not it has been articulated with this particular vocabulary, a goal of feminism has been to make visible the ubiquitous cultural, political, social, and economic infrastructures and the roles of women within them. While infrastructures are usually transparent, the structures created within them can be more consciously designed. We can understand "infrastructure" to indicate a large type of immersive and network-like system. The Latin prefix "infra-" means that which is below the surface or foundational, and "structura" relates to the process of building or construction. As "structures below the surface," infrastructures may be of such a large scale that they are difficult to understand or grasp as a whole and cannot be easily mapped. They are not planned out in their entirety with a singular purpose, and they often cannot even be pointed to physically. In contrast, the word "structure" describes a smaller part of an infrastructure—one built, designed, organized, or curated purposefully and visibly. Databases can be understood as types of data structures. Of particular importance for us now that informational infrastructures have become

globalized are the structures that collect and store data. Popular web applications—from Wikipedia to Pinterest to Facebook—are built upon huge databases. The content of these sites often come under scrutiny—for example, activist groups have attempted to address and correct the ways that women are underrepresented on Wikipedia's pages (Wadewitz, 2013). Even beyond questions of content, however, we might ask how the underlying classification and organizational schema themselves might be gender-biased. We could also look at how the categories residing in data structures perpetuate Western-centric values. Because data structures are shared worldwide, we need to consider questions of privilege and power within them.

Addressing Gendered Standards and Classifications

The gender problem within data classification systems has been around for a long time. Working from the field of library information systems, Hope Olson (2001) describes how the original architects of library classification systems decided on organizational schemes. Charles Cutter, who published *Rules for a Dictionary Catalogue* in 1904, advised for uniformity of categories except in cases where it would be more convenient for "the" public to have things listed in a non-uniform way. Olson argues that his language indicates a belief in a singular public whose members all share the same worldview; in other words, "a universality is present in Cutter's view, but it is the singular public who defines it" (p. 642). Problematically, Cutter's "singular public" is inclusive of all community members, but rather, it is "a particular part of humanity that shares cultural, social, or political interests. That idealized community excludes individuals and groups who do not share its interests" (Olson, 2001, p. 643). Some of the earliest cataloguing systems, upon which much of current practices are based, privilege hierarchical relationships; broader terms channel narrower terms underneath them (p. 644-645). Sub-categories are not evenly distributed and favor a male-privileged worldview. Olson gives the following example to illustrate how this happens:

> The subdivision "- Relations with women" subtly reinforces the subject/object roles of men and women. There is no parallel under "Men" (one cannot express Simone de Beauvoir's relations with men as one can express Jean-Paul Sartre's relations with women). This anomaly reflects mainstream culture's positioning of men as knowing subjects in our society and women as objects to be known, the objects of men's relationships. (p. 647)

Another result of this categorization system is that works "embodying multiple marginalizations" are "either ghettoized in an obscure corner of the catalog (all women

or all African Americans lumped together) or dispersed in a diaspora of little ghettos. Separated from mainstream subject classifications, where they are pushed to the margins, they will not disturb library users looking for books on 'real' topics" (p. 658-659).

Olson does not stop at critique, however. Instead, she looks for alternative systems of organization and ways of searching for library information that would avoid the problems of marginalization and ghettoization, often due to hierarchical classification structures. She contemplates the benefits and problems that come with "free text searching"—remarking that this strategy could be useful in finding "topics not representable in a controlled vocabulary" but would also return too many results (p. 660). Another suggestion is to use alternative names so that, for example, a search for "wimmin" would return the same values as a search for "women" without instructing users to search again for "women." Yet another possibility would be to use past transaction logs to aid with current searches (p. 661). She emphasizes that change will come only when women work to modify the already established systems. By suggesting these alternative search structures, Olson argues that cataloguing systems should stop assuming that there will be just one type of user who represents a singular public. Library catalogues need to relinquish some of their structural power to users of all identities. Such a hypothetical catalogue would communicate ideas of inclusivity and equality.

Olson's analysis of library cataloguing systems provides just one example of how we might think about reorganizing data structures to reflect gender and race equality. Another example of efforts to reframe women's historical writings in feminist terms can be found in a feminist-oriented, curated data structure: the *Orlando* database.

Feminist Databasing

Published online by Cambridge University Press in 2006, *Orlando: Women's Writing in the British Isles from the Beginnings to the Present,* provides information on 1,300 women writers and bibliographic references on over 25,000 titles. It does not provide the texts themselves, but it does provide "new biographical and critical accounts of the lives and works of its subjects, together with contextual materials relevant to critical and historical readings" (Brown, Clements, & Grundy, 2006). It was created and edited by a team of three women, Susan Brown, Patricia Clements, and Isobel Grundy, along with a large team of co-investigators, technical personnel, research associates, post-doctoral fellows, and research assistants (Brown et al, 2006-2015). According to the

scholarly background webpage, efforts towards recovering the work of women writers have been underway as the "Orlando Project" since the 1960s. As the site conveys, "This phenomenally vigorous scholarly work of inclusion—of writers omitted from traditional historical accounts, at least partly by reason of gender or race or class—is arguably the major feature of recent literary historical scholarship" (Brown et al, 2006-2015). Within the scholarly introduction, the following describes how Orlando positions itself and its purpose:

> *Orlando focuses on gender, and it emphasizes the intellectual, material, political, and social conditions (including writing by men) that have, over time, helped to shape writing by women. It sees gender as an indispensable tool for historical analysis that helps to shape the questions we ask about the production, reception, and features of written texts and about the ways in which these have been understood throughout the history of women's writing. (Brown et al, 2006-2015)*

Here, the editors of *Orlando* convey its basic rhetorical context, audience, and purpose: it arises out of a need to recover women's contributions to literary history, it seeks to emphasize the conditions that have shaped women's writing and reframe historical analysis, and its primary audience appears to be literary critics and scholars. While it focuses on "literature," the database does include "women known as writers of science, household advice, or popular genres, and those known (if at all) mostly for non-literary reasons who also left significant writing" and some male writers who provide textuality (Brown et al, 2006-2015). Looking at what the database intends to communicate through its stated purposes, however, provides only one level of understanding. Looking at how the data are sorted and classified garners a more thorough analysis.

Orlando is organized not hierarchically, but through a system of tagging. The editors fully realize that this system of tagging is highly interpretive, based on what the historians, as architects of the system, prefer to prioritize and communicate as most important. They note this realization in their writings about the process of tagging during the project (Brown, Grundy, Clements, Elio, Balazs, & Cameron, 2004; Butler, Fisher, Coulombe, Clements, Grundy, Brown, … & Cameron, 2000). Butler et al. (2000) explain that their work does not involve applying tags to existing texts. Rather, they tag the descriptive histories that they compose in the database. Using SGML, they create three distinct document types (DTDs): biography, writing, and events. They model these structurally after the Text Encoding Initiative (TEI), adding interpretive tags as they see fit. They write,

> *For example, the biography DTD has tags for birth, family, education, and political affiliations; writing documents use tags for such specific information as genre, intertextuality, literary awards, and relations with publishers; events documents contain chronological events that have such information as organization names and places tagged. (Butler et al, 2000, p. 112)*

As Butler et al. describe, the process of applying tags to their interpretive histories is complex and problematic. Because they have so many different people working on tagging, many of them postdoctoral students who average a little more than one year working on the project, it is nearly impossible to achieve consistency. They report that as of 2000, there were 238 "unique element types" in their DTDs and 230 "unique attributes." The process of deciding on criteria is described as collaborative—"we had the sense of a shared common understanding of what each tag and attribute meant" (p. 112). They provide an example of one DTD element, "political affiliation," that encapsulates and documents the process of creating it and testing it (p. 113). However, they encountered a need to edit for consistency among variables and automated this process using a database. They found that beyond core attributes such as names and places, it was often extremely difficult to systematically manage various tags.

The *Orlando* editors offer considerable reflection on the tagging process, but they do not offer an extended discussion of why or how certain terms were chosen and applied. For example, they do not offer an explanation of the possible genres that works have been assigned or the thought process behind assigning them. There does not appear to be any reflection on the ways that genre can be rhetorical or reflective of a particular worldview, or as a type of social action (see Miller, 1984). A group of literary scholars compose the non-core tags and lesser attributes according to a sort of *folksonomy*—a term coined in 2004 by Thomas Vander Wal to describe how everyday users of websites (for example, Flickr or Blogger) tag content according to their own associations and definitions rather than relying on traditional hierarchical taxonomies. This approach to classification echoes the one suggested by Hope Olson as a remedy for marginalization and ghettoization in libraries. Yet, *Orlando*'s tags are not purely folksonomic—tags are "cleaned up" via automated database algorithms, and taggers must go through a training session where they are taught specific protocols to follow. Despite efforts to consciously structure data based on the knowledge, input, designs and expectations of its collaborators, many of the classifications and standards that organize the *Orlando* database remain invisible. In the case of genre identification, the taggers must share a common understanding of genre. There are parts of the tagging system that are consciously articulated as standardized—Figures 1, 2, and 3 serve as examples, but

there are other tag diagrams as well. The total of all the mapped nodes do not represent all possible tags within the database. Other classifications are left up to the discretion of the tagger, presuming a shared epistemology or knowledge infrastructure, as in the genre example.
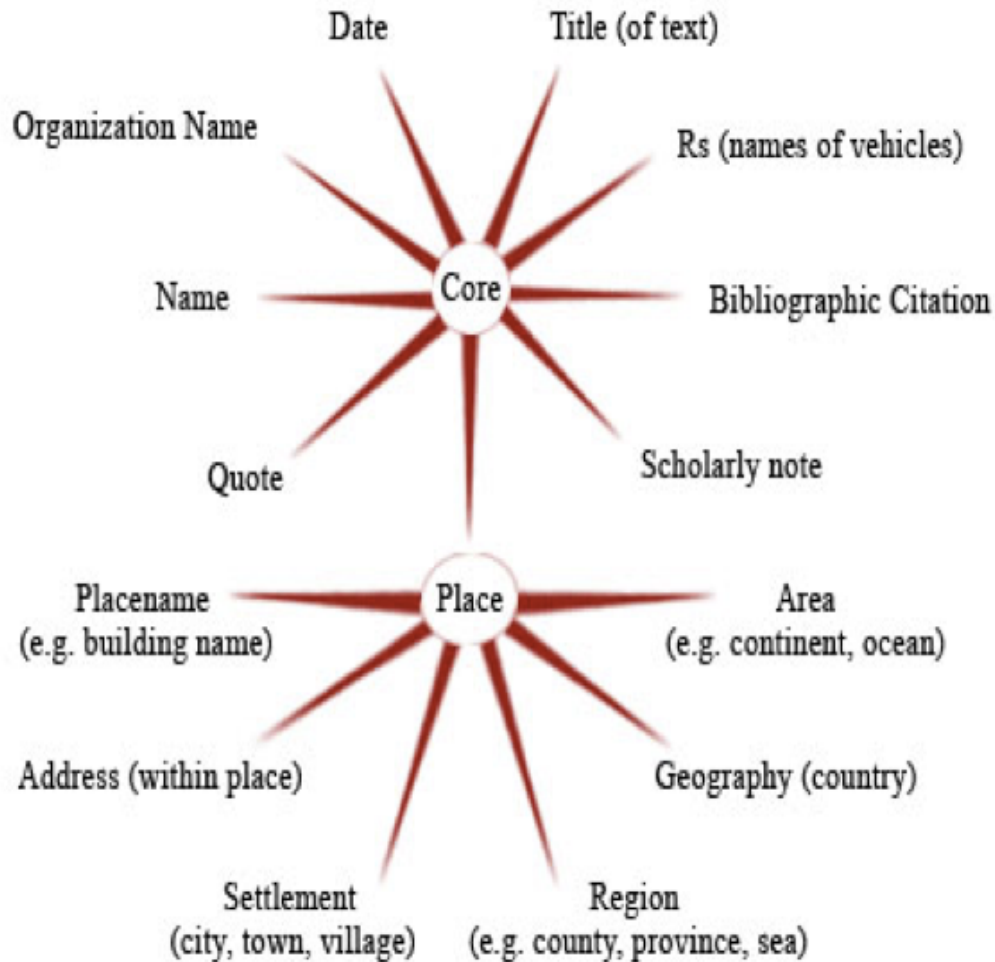
(https://adanewmedia.org/wp-content/uploads/2015/10/masters_figure1.png)

Figure 1. Orlando Database Core Tags (Brown et al, 2006)
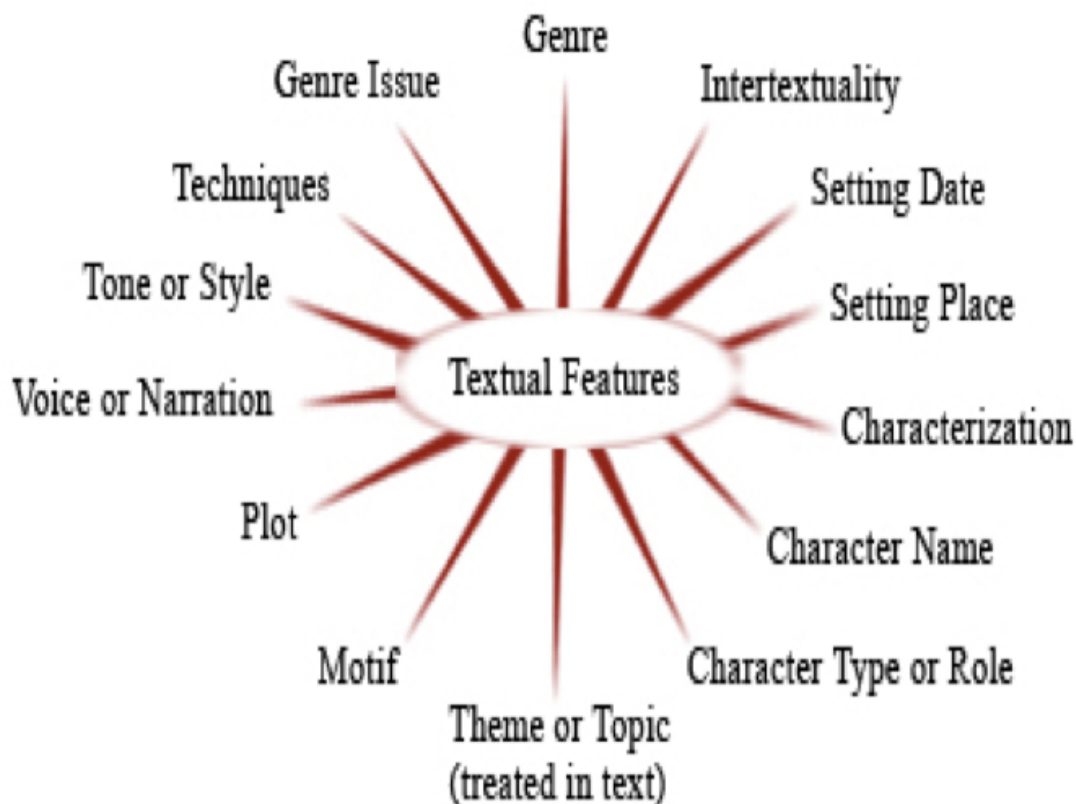
(https://adanewmedia.org/wp-content/uploads/2015/10/masters_figure2.png)

Figure 2: Orlando Database Textual Features Tags (Brown et al, 2006)

(https://adanewmedia.org/wp-content/uploads/2015/10/masters_figure3.png)

Figure 3: Orlando Database Life Tags (Brown et al, 2006)

The consciously articulated purpose of this data structure is to facilitate easy searches on the part of the user. However, there is a purpose not articulated by the editors, probably because it is unconscious—editors seek to reinforce existing knowledge infrastructures within the culture and society of literary scholarship. These tag diagrams, provided as keys within the pages of *Orlando,* allow viewers to click on individual terms and view descriptions for each attribute. For example, if we click on "Cultural Formation" within the "Life" tag diagram, a listing appears that contains a definition of the term, related tags, and examples. In this case, "Cultural Formation" has two sub-elements: 1) discursive accounts of "class issue, nationality, issue, race and ethnicity, religion, sexuality"; and 2) an additional level of tagging to define identity based on "race, colour, class, national heritage, nationality, geographical heritage, ethnicity, denomination, language (within cultural formation), political affiliation, and sexual identity" (Figure 4). Further clicking on the terms for these levels reveals more information but not a complete list of options for labeling. Taggers would assign labels as they see fit, based on their knowledge of the author and their literary works. In this way, links between the consciously designed structure of the *Orlando* database and larger infrastructures of the literary community become created and reinforced. Because this infrastructural work is historical—it involves mapping out knowledge of the past—the taggers must be interpretive and reprocess already collected data. The *Orlando* editors recover women's writings and establish validity by building a knowledge framework around it. In order to persuade an audience that this knowledge is valid, the editors use already existing standards and classifications that have currency in larger literary or cultural circles. By using a system of tags that are already familiar to literary scholars, *Orlando* legitimizes women's history by fitting it into an existing knowledge framework that has been traditionally male-centered. This process accomplishes important feminist recovery work, promoting awareness of women writers within a traditionally male-dominated cultural infrastructure. At the same time, it raises questions about the data-structuring process itself. Can there be specifically feminist ways of working with data? Can there be such a thing as a feminist data structure?

(https://adanewmedia.org/wp-content/uploads/2015/10/masters_figure4.png)

Figure 4: Cultural Formation Element Description (Brown et al, 2006)

Conclusion: Conscious Structuring

The word "structure" holds similar connotations to the word "system." In a recent interview for "DOCC 2013: Dialogues on Feminism and Technology," Lucy Suchman and Katherine Gibson discuss the intersections of feminism, technology, and systems (Balsamo, Suchman, & Gibson, 2013). Suchman proposes that the term "system" itself has a modernist, rationalist association to which she feels ambivalent. Gibson agrees on this connotation and offers that an alternative to emphasizing systems would be to emphasize how things exist in relation to one another. Gibson emphasizes that the term "economic system" has been used as a master signifier to describe one mode of economics—capitalism—as the dominant economic reality against which every other type of economic activity always gets positioned. In actuality, she explains, there are

many other forms of economic activities and relations that are pervasive but that have not been focused on extensively, many often associated with women, for example, gift-giving and reciprocal economic activities that involve spheres of production and reproduction. She has become increasingly interested in relational types of thinking that eschew a belief in one dominant "system" (Balsamo et al, 2013). While both words are interchangeable to a great extent, "infrastructure" holds more open and relational connotations than the word "system" because "infrastructure" connotes a web of interdependencies that are contingent upon relationships and that build with practices over time. Systems are in fact infrastructures; yet when we think there is only one overarching primary system, we tread into dangerous waters. If someone claims that a system is somehow "natural" without attempting to invert it and see how it arises through a multitude of dependencies, red flags should go up. What we think of as "natural" is also transparent and infrastructural. If a structure is to reflect feminist principles, then, it should work towards being non-transparent—in the sense that it is outlined and viewable—yet transparent in the sense that it does not hide its motivations.

Ultimately, a feminist data structure might take cues from what Jo Freeman (aka Joreen) advocates in "The Tyranny of Structurelessness" (1970-1973). Writing about group organization within the feminist movement, Joreen notices that the ideal of "structurelessness" does not work; a few "informal elites" always end up directing what happens unless a group adopts principles of democratic structuring. If we carry this line of thinking into the realm of organizing data, feminist data structure would be one where classification categories are consciously articulated and decided as fairly as possible by those who will access or interact with it, not just by an elite few. The recent Feminist Wikipedia movement follows this model. For literature and writing scholars, feminist structuring might mean not taking categories within genre classifications as given or natural, especially when genres have arisen within historically Western and male-dominated literary contexts. Assumptions about categories should not be taken for granted, but constantly questioned. Feminist data structuring processes, with equality as the goal, would involve a great amount of reflection, articulation, and collaboration. Ideally, these strategies could apply to address racial and global marginalizations within data structures.

References

Balsamo, A., Suchman, L., & Graham, K. G. (2013). Feminism, technology, and systems 2: Infrastructures. *FemTechNet*. Retrieved from http://vimeo.com/79740274#at=0

(http://vimeo.com/79740274#at=0) .

Bowker, G. (2005). *Memory practices in the sciences.* Cambridge, MA: MIT Press.

Bowker, G. & Star, S. L. (1999). *Sorting things out: Classification and its consequences.* Cambridge, MA: MIT Press.

Brown, S., Grundy, I., Clements, P., Elio, R., Balazs, S., & Cameron, R. (2004). Intertextual encoding in the writing of women's literary history. *Computers and the Humanities 38,* 191-206.

Brown, S., Clements, P., & Grundy, I. (The Orlando Project) (2006-2015). *Orlando: Women's writing in the British Isles from the beginnings to the present.* Cambridge: Cambridge University Press. Retrieved from: **http://orlando.cambridge.org/** (http://orlando.cambridge.org/) .

Butler, T., Fisher, S., Coulombe, G., Clements, P., Grundy, I., Brown, S., … & Cameron, R. (2000). Can a team tag consistently?: Experiences on the Orlando project. *Markup Languages, 2*(2), 111-125.

Freeman, J. (1970-1973). The tyranny of structurelessness. In *JoFreeman.com (feminist articles by Joreen).* Retrieved from **http://jofreeman.com/joreen/tyranny.htm** (http://jofreeman.com/joreen/tyranny.htm) .

Miller, C. (1984). Genre as social action. *Quarterly Journal of Speech 70,* 151-167.

Olson, H. (2001). The power to name: Representation in library catalogs. *Signs 26.3,* 639-668.

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.

Wadewitz, A. (2013, July 26). Wikipedia's gender gap and the complicated reality of systemic gender bias. In *HASTAC.org (HASTAC scholars blog posts).* Retrieved from **http://www.hastac.org/blogs/wadewitz/2013/07/26/wikipedias-gender-gap-and-complicated-reality-systemic-gender-bias** (http://www.hastac.org/blogs/wadewitz/2013/07/26/wikipedias-gender-gap-and-complicated-reality-systemic-gender-bias)

‹ **DATA**   ‹ **INFRASTRUCTURE**   ‹ **PEER REVIEWED**   ‹ **TAGGING**   ‹ **VISUALIZATION**
‹ **WOMEN IN LITERATURE**

Christine L. Masters (https://adanewmedia.org/author/christinemastersjach)
Christine L. Masters is a doctoral candidate in English at Purdue University. Her primary concentration is rhetoric and composition, with secondary areas in digital rhetorics and professional and technical communication. Her dissertation explores how new materialist feminisms could further impact understandings of writing and learning, as well as inform a new approach to teaching composition that involves analyzing and creating data structures.

2 THOUGHTS ON "WOMEN'S WAYS OF STRUCTURING DATA"

Pingback: **Now live: Ada Issue no.8, Globalization, Gender and the Digital I**

Pingback: **Digital Humanities & Data Science: A Short Reading List | Cindy A. Nguyen**