

The Trouble with Genre

Brian S. Matzke

University of Michigan School of Information, Ann Arbor, Michigan, USA



Abstract:

Jacques Derrida famously asked in “The Law of Genre” (1980), “Can one identify a work of art...if it does not bear the mark of a genre?” As metadata elements, genres are both highly useful and highly unstable; the field lacks consistent and widely agreed-upon definitions of genres, and texts can evince traits of many genres in different ways and to different degrees. Based on a preliminary survey of recent literature, this poster identifies four common issues around the use of genre in digital libraries and identifies potential avenues for future research.

Distant Reading

	Close Reading	Distant Reading
What?	Detailed language of <i>individual</i> text	Whole range of genre, period, or national literature
How?	Qualitative methods, individual analysis	Quantitative methods, reader surveys, digital libraries
Literature as Metadata About:	Human experiences	Human cultures

- ◆ “Close” and “distant” reading practices are not simply literary analyses taking place at different scales, but rather different understandings of the role of scale in literary analysis (Jin 2017). The methods are complementary, but the fundamental assumptions about what is being gleaned through scholarly analysis are often quite different.
- ◆ While “distant” is often associated with digital humanities (DH), it is important to remember that DH methods can also be employed in close-reading-type projects, e.g., network analyses of characters in a single novel (English and Underwood 2016). It is also important to note that “analog” humanists likewise organize their work into categories, be that genre, period, region, etc.
- ◆ When it comes to data, big isn’t always better (boyd and Crawford 2011). DH methods may also be applied to “Mid-Range Reading”; i.e., analyses of smaller corpora (Booth 2017).

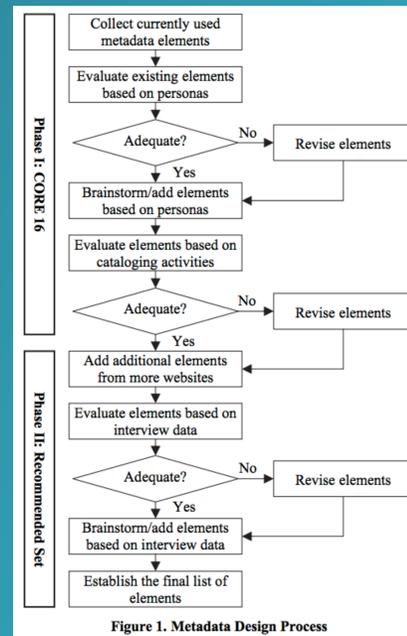


Figure 1. Metadata Design Process

Experts vs. Lay Users

When constructing a genre schema, who gets to define the terms? Relying solely on the expertise of literary critics or librarians risks excluding the valuable perspectives of everyday people and marginalized voices, while relying on lay users risks imprecision and inconsistency in the application of genre terms (e.g. Wilson 1968).

- ◆ Lee, Cho, Fox, and Perti’s (2013) model for an iterative process (pictured here) shows how user input can contribute to the construction of a metadata schema. This has the advantage of rigorously capturing users’ perspectives, but has the disadvantage of being so particular to the collection that federating with other repositories would be difficult.
- ◆ Social tagging provides an opportunity to capture users’ knowledge about genre, but studies of social tagging have been inconsistent, sometimes showing significant overlap with LCSH terms (Yi and Chan 2009), and other times showing a great deal of divergence (Lu, Park, and Hu 2010). Intercoder reliability remains a major concern with social tags—garbage in, garbage out.
- ◆ Companies like Netflix and Amazon may be useful resources, but their profit motives may shape their metadata in ways that merit critical pressure (English 2016).

The Vocabulary Problem



“Genre has long been a source of uncertainty and unease in bibliographic control.... Genre is, in essence, an integration of aboutness, of-ness, and is-ness for its multidimensions”
- Zhang and Olson 2016

The way we talk about and use genre is full of ambiguity. Thus, creating a controlled vocabulary that provides consistent, accurate, and precise values with which to populate a metadata field is a daunting task.

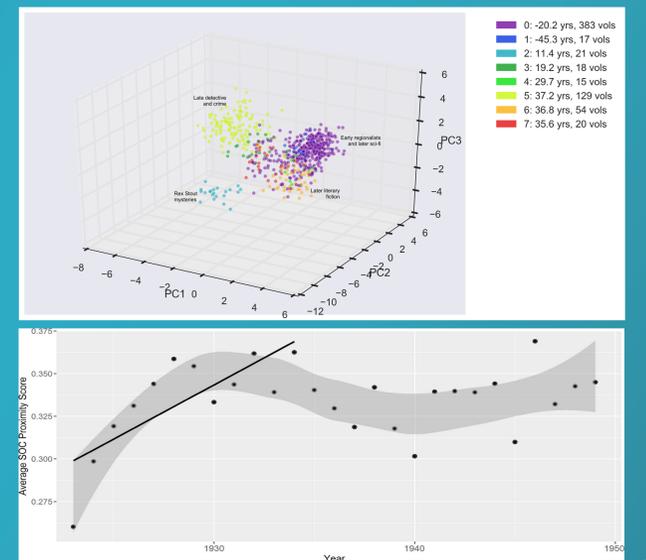
- ◆ The Library of Congress Subject Headings (LCSH) vocabulary is insufficient for many research needs. Social tagging provides one model for compensating for deficits in LCSH, but social tagging has its own limitations (Murphy and Rafferty 2015).
- ◆ Ambiguity increases with the level of granularity in genre terminology. Distinguishing between poetry and prose or between fiction and nonfiction is relatively straightforward; distinguishing between realism and modernism or between horror and science fiction is a more challenging enterprise (Green 2014).
- ◆ Netflix’s database employs a syntax that combines descriptive phrases into 76,897 different “microgenres”; such strings of phrases may provide a model for future work on genre classification (Madrigal 2014).

Machine Learning

“Machine learning...represents a new way of thinking about literary concepts, like genre, that may be organized around loose family resemblances rather than crisp definitions”
- Underwood 2017

Scholars have recently been drawn to opportunities created by network analysis and computational stylistics to “discover” genres, i.e., generate clusters of texts based on previously invisible affinities. This ostensibly represents a way out of the problems of bibliographic control and controlled vocabularies.

- ◆ Machine learning may provide ways to define nebulous cultural distinctions, e.g. between “literary fiction” and “genre fiction”, or to critique such distinctions, as in Wilkens’s (2016) study (pictured above).
- ◆ It is doubtful that the generic affinities unearthed through machine learning will have any utility without domain knowledge to provide context, and it can be productive to test algorithms against expert knowledge, as in Long and So’s (2016) analysis of the Stream of Consciousness novel (pictured below).



References

Booth, Alison (2017). Mid-Range Reading: Not A Manifesto. PMLA 132(3): 620-627.
 boyd, danah and Crawford, Kate (2011). Six Provocations for Big Data. A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society.
 English, James F. (2016). Prestige, Pleasure, and the Data of Cultural Preference: 'Quality Signals' in the Age of Superabundance. Western Humanities Review 70(3): 119-139.
 English, James F. and Ted Underwood (2016). Shifting Scales. Modern Language Quarterly 77(3): 277-295.
 Green, Harriet E. (2014). Literature as a Network. portal: Libraries and the Academy 14(2).
 Jin, Jay (2017). Problems of scale in “Close” and “Distant” reading. Philological Quarterly 96(1).
 Lee, J. H., H. Cho, V. Fox, A. and Perti (2013). User-centered approach in creating a metadata schema for video games and interactive media. Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries.
 Long, Hoyt and Richard Jean So (2016). Turbulent Flow: A Computational Model of World Literature. Modern Language Quarterly: A Journal of Literary History 77(3): 345-367.
 Lu, Caimei, Jung-ran Park, and Xiaohua Hu (2010). User tags versus expert-assigned subject terms: A comparison of LibraryThing tags and Library of Congress Subject Headings. Journal of Information Science 36(6).
 Madrigal, Alexis (2014). How Netflix Reverse Engineered Hollywood. The Atlantic.
 Murphy, Helen, and Pauline Rafferty (2015). Is there nothing outside the tags?: Towards a poststructuralist analysis of social tagging. Journal of Documentation 71(3): 477-502.
 Underwood, Ted (2017). A Genealogy of Distant Reading. DHQ: Digital Humanities Quarterly 11(2).
 Wilkens, Matthew (2016). Genre, Computation, and the Varieties of Twentieth-Century U.S. Fiction. CA: Journal of Cultural Analytics.
 Yi, Kwan, and Lois Mai Chan (2009). Linking folksonomy to Library of Congress subject headings: an exploratory study. Journal of Documentation 65(6): 872-900.
 Zhang, Lei, and Hope A Olson (2016). Distilling Abstractions: Genre Redefining Essence versus Context. Library Trends 63(3): 540-554.