

---

# Cultural (Re-)formations: Structuring a Linked Data Ontology for Intersectional Identities

**Susan Brown**

sbrown@uoguelph.ca  
University of Guelph, Canada

**Abigel Lemak**

alemak@uoguelph.ca  
University of Guelph, Canada

**Colin Faulkner**

cfaulk01@mail.uoguelph.ca  
University of Guelph, Canada

**Kim Martin**

author.email@domain.com  
University of Guelph, Canada

**Rob Warren**

rwarren@math.carleton.ca  
Carleton University, Canada

---

## Introduction

Cultural diversity has been an increasing source of debate within the digital humanities community. The concentration within the *Debates in Digital Humanities* series (Gold, 2012; Gold and Klein, 2016) of pieces reflecting the increasing prominence of matters related to race, gender, cultural diversity and difference is but one marker of the extent to which diversity matters. The Orlando Project in feminist literary history incorporated an intersectional understanding of identity categories from the outset (Brown, Clements and Grundy, 2006-2017). Translating Orlando's Extensible Markup Language (XML) data into linked open data (LOD) to make it accessible, interoperable, and amenable to a range of analytical approaches (Simpson and Brown) requires an ontology that will serve both Orlando and the broader research community hosted by the Canadian Writing Research Collaboratory (CWRC). This paper outlines the CWRC ontology design and the

challenges of shifting from semi-structured to structured data (Smith, 2016: 273).

Much work on digital diversity expresses skepticism of the ability of systematized knowledge structures to capture the performative, processual, and contingent nature of lived subjectivities. Tara McPherson stresses that "computers are themselves encoders of culture" and calls for more attention to be paid to the interconnectedness of the structures of code and the management of race socially: "Just as the relational database works by normalizing data—that is, by stripping it of meaningful, idiosyncratic context, creating a system of interchangeable equivalencies—our own scholarly practices tend to exist in relatively hermetically sealed boxes or nodes." Scholars including Lisa Nakamura (2002: 120) and Moya Bailey (2011) see value in "messiness" as a way to push against and redefine the contours of a digital humanities scholarship that remains rooted in predominantly white epistemology.

At the same time, relegating representations of difference to narrative rather than structured data will produce gaps within big data that are both impoverishing for humanities inquiry and dangerous in their political implications (Lerman, 2013; Trevinarus, 2014; "Use"; Brown and Simpson, 2013). Adriel Dean-Hall and Robert Warren (2013) have advocated approaches that respect the privacy and preferences of lived human subjects while improving the responsiveness of online systems to diversity and complexity. Within a LOD context, what are finally findable, processable, and reusable on the global graph are things, not strings, so the challenge is the extent to which nuance, context, and indeed messiness can be incorporated into a LOD ontology.

The Orlando Project (Brown, et al., 2006-2017) charted a middle ground between narrative and structure for its bespoke XML tagset. The team struggled with the hierarchical nature of XML particularly in relation to identity categories, torn between knowledge that readers would turn to Orlando to find writers associated with particular cultural identities and recognition that such categories are discursive rather than essential (Fuss, 2013). It devised a "Cultural Formation" tagset to depict identity as neither unitary nor immutable, and as much related to representational acts as to the lived experiences into which those representations blur. Precisely because constituted through discursive and social practices, vocabularies associated with subjectivities and identities can shift over time and place, and throughout an individual's lifetime.

## Cultural formation tagset

The Cultural Formation (CF) tagset recognizes categorization as endemic to social experience, while incorporating variation in terminology and contextualization of identity categories by employing tags at different discursive levels. CF tags describe the subject positions of individuals through 1) contextual tags that encode substantial discussions: class; language; nationality; race and ethnicity; religion; and sexuality; and 2) granular tags that describe, in a word or short phrase, class; ethnicity; gender; geographical heritage; language; nationality; national heritage; political affiliation; race or colour; religious denomination, and sexual identity. With the exception of gender and social class, the Orlando schema eschewed fixed attribute values for the granular tags, allowing the prose to employ the most appropriate language for the context. The structure was not entirely logical or parallel, and we are making the ontology more consistent. The granular tags possess attributes regarding forebears and whether a subject self-identified with a particular term. The tagset aimed to highlight the extent to which social classification is culturally produced and discursively embedded. Rather than disambiguating leaky cultural categories, it considered them as mutually constitutive with historically specific discursive frameworks, including our tagging structures.

CF encoding pointed users towards a framework for raising and debating complex matters for cultural investigation rather than standardized classifications, refusing to neatly group writers into distinct and fixed categories, since those categories were neither stable nor mutually exclusive (Algee-Hewitt, Porter, Walsler, forthcoming). It can represent quite complex identities, as in the case of Anna Leonowens, the writer whose story of life as governess to the royal Siamese harem was popularized in *The King and I*. Partial markup for the first paragraph of her CF description is shown in Figure 1.

Although AL herself, in attempting to adopt an unequivocally `< NATIONALITY SELF-DEFINED=SELFYES > English </NATIONALITY>` identity, implicitly claimed that she was `< RACECOLOUR SELF-DEFINED=SELFYES > white </RACECOLOUR>`, evidence suggests that while her father was probably `< NATIONALHERITAGE > Welsh </NATIONALHERITAGE>` (he had lived in `< PLACE > < REGION > Middlesex </REGION> < GEOG REG=England > </GEOG> </PLACE>` ) and presumably white, her mother was quite possibly `< RACECOLOUR > Eurasian </RACECOLOUR>`. [citations omitted] If this is the case then AL suppressed her mixed-race origins.

Figure 1: Adapted from Brown, Clements and Grundy, "Anna Leonowens", Life tab, Show Markup option

The CF component of Orlando's knowledge representation is thus crucial to its intersectional approach

to identity (Brown et al., 2006). Creating a LOD ontology that was not self-referential, however, requires translating the strings or literal values from CF tags, to link Orlando's semantic structures to other semantic web communities.

## LOD ontology creation

An ontology "is a formal naming and definition of the types, properties, and interrelationships of the entities that really or fundamentally exist for a particular domain of discourse" (Wikipedia, Ontology - Information Science). Using a standard ontology language such as OWL allows others to interact and exchange with a particular view of the world through a computational process of mediation. As a representation of that understanding, an ontology can be referenced, (dis)agreed with, extended, and used operationally. The coexistence of different representations provides the foundation for translations between LOD concepts.

Ontology creation in our case, as in many others, was driven by the idiosyncrasies and limitations of an existing data set. The information architectures of application databases or XML stores are not always reconcilable to a consistent information system. The CF tagset represents a major challenge in that its structure was designed to eschew disambiguation. Even the major tags were difficult to relate within a concise ontology (Figure 2).

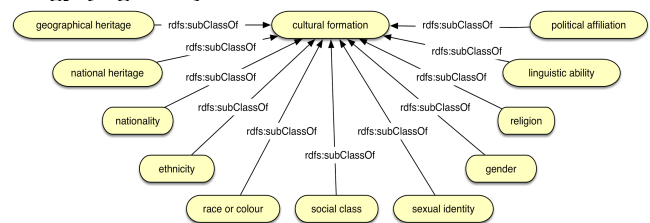


Figure 2: Schematic representation of the granular Cultural Formation tags from Orlando (Please note that these representations are simplified in order to make them legible to the reader.)

For example, nationality and national heritage are not employed as commensurate with citizenship, a well-defined legal concept related to an organized state. They can also be related to a geographical area, which may or may not coincide with a state. Finally, nationhood can reference socio-political constructs such as Lesbian Nation (Johnston, 1973; Ross, 1995; Munt, 1998) or disavowals of nationality such as Virginia Woolf's (1938: 197), which Orlando quotes alongside assigning Woolf an English nationality, a contradiction that requires contextual evidence to make sense.

## Linked into context

We decided to make all human-readable annotations within the dataset instances of contextual notes to which the ontological classes are directly tied (Figure 3).

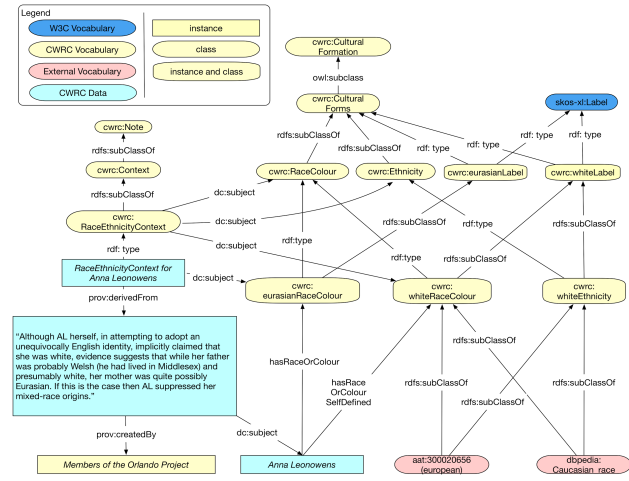


Figure 3: Schematic representation of how the discursive context (note) links to the classificatory structure, and how classificatory labels relate to predicates and external ontologies. Skos:narrower/broader relationships are also used, but omitted here to improve legibility

Thus we model the discursive context within a Race[or]EthnicityContext class. The note instance links to instances of granular category labels, here RaceColour; it provides the provenance and the basis for links to source information. Linking to the provenance of the LOD is particularly important for disputed or contradictory information, as in our example. We are modeling the original Orlando narrative as a source document for our LOD provenance using the the Web Annotation Data Model’s subproperty instances. We aim to link every triple to the prose from which it is derived, providing provenance information and contains citations to the sources on which identity assertions are based.

## Relating cultural formations

Cultural formation for Orlando is understood primarily as representational, which is not to say that cultural formation is not real or that it has no material effects. The complex signifiers of cultural identities float across Orlando tags as cdata or free text in a semi-structured representation of cultural identities and categories. For the CWRC ontology, we strategized to relate this ontological perspective to that of external

vocabularies without conflating our truth with theirs. Our architecture does not import other ontologies wholesale, but adopts components of major vocabularies such as BIBO, FOAF, and FRBR, and relates to large vocabularies in defined ways. As indicated in Figure 3, the instances of cwrc:whiteRaceColour and cwrc:whiteEthnicity within the CWRC ontology are subclasses of the cwrc:whiteLabel. This retains the ambiguity of terms such as “white” or “Jewish” precisely as labels that draw together particular types of identity categories, as well as subclasses of those labels. As indicated, those subclasses can be linked to terms in external vocabularies, but both internal and external terms are understood within the CWRC ontology as labels. Indeed, constructing this ontology has brought home to us the need for the LOD community to think through with greater care the relationship between representation and “reality” in LOD ontologies. A further complication is that identity categories are not only historically contingent but often also change over a particular individual’s lifetime. The Orlando dataset supports such nuance in only a few cases, so we have not started with this gnarly problem, but we aim to build into the ontology the capacity to represent such cultural formation dynamics in order to accommodate more temporally precise data.

## Conclusion

The CWRC ontology design avoids representing RDF extractions from Orlando data as positivist assertions, and yet produces machine-readable OWL/RDF-compliant graph structures. It allows references to, without endorsing, external ontological vocabularies that are nevertheless part of documenting intersectional cultural processes and identities.

We will present CWRC ontology as built around the CF design described here, and we will demonstrate its implications through several practical examples. Figure 4 shows schematically the intersectionality of multiple identity categories associated with Leonowens, including the ways that instances are related by subclass relationships in accordance with OWL principles. This importantly allows us to reference components of other ontologies (here the Muninn Appearances ontology, Library of Congress Subject Headings, Getty Art and Architecture Thesaurus, and DBpedia) without adopting them wholesale.

[Schema/blob/master/schemas/orlando\\_biography.rng](http://Schema/blob/master/schemas/orlando_biography.rng)

## Bibliography

Alexiev, V., Cobb, J., Garcia, G., and Harpring, P. (2016). *Getty Art and Architecture Thesaurus*. J. Paul Getty Trust. <http://vocab.getty.edu/doc/queries>

Algee-Hewitt, M., Porter, J. D. and Walser, H. (Forthcoming, 2017). "Representing race and ethnicity in American fiction: 1789-1964."

Bailey, M.Z. (2011). "All the digital humanists are white, all the nerds are men, but some of us are brave." *Journal of Digital Humanities* 1.1. <http://journalofdigitalhumanities.org/1-1/all-the-digital-humanists-are-white-all-the-nerds-are-men-but-some-of-us-are-brave-by-moyaz-bailey/> (accessed 7 April 2017)

Brickley, D., and Miller, L. (2000-2014). *FOAF Vocabulary Specification 0.99*. <http://xmlns.com/foaf/spec/>

Brown, S., Clements, P., and Grundy, I (eds.) (2006-2017). *Orlando: Women's Writing in the British Isles from the Beginnings to the Present*. Cambridge: Cambridge University Press Online.

Brown, S., Clements, P., and Grundy, I. (2006). "Sorting things in: Feminist knowledge representation and changing modes of scholarly production." *Women's Studies International Forum* 29.3.

Brown, S., & Simpson, J. (2013, October). The curious identity of Michael Field and its implications for humanities research with the semantic web. In *Big Data, 2013 IEEE International Conference on* (pp. 77-85). IEEE.

Canadian Writing Research Collaboratory. (n.d.) <http://cwrc.ca>

D'Arcus, B., and Giasson, F. (2008-2013). *Bibliographic Ontology Specification (BIBO)*. <http://purl.org/ontology/bibo/> Structured Dynamics.

Davis, I., and Newman, R. (2005). *Functional Requirement for Bibliographic Records (FRBR)* <http://purl.org/vocab/frbr/core#>

DBpedia. (n.d.) <http://wiki.dbpedia.org/>

Dean-Hall, A. and Warren, R. (2013). "Sex, privacy, and ontologies." *SEXI*. Rome, Italy. <http://www.dbdump.org/~warren/publications/dean-hall:sexi:2013/dean-hall:sexi:2013.pdf> (accessed 7 April 2017).

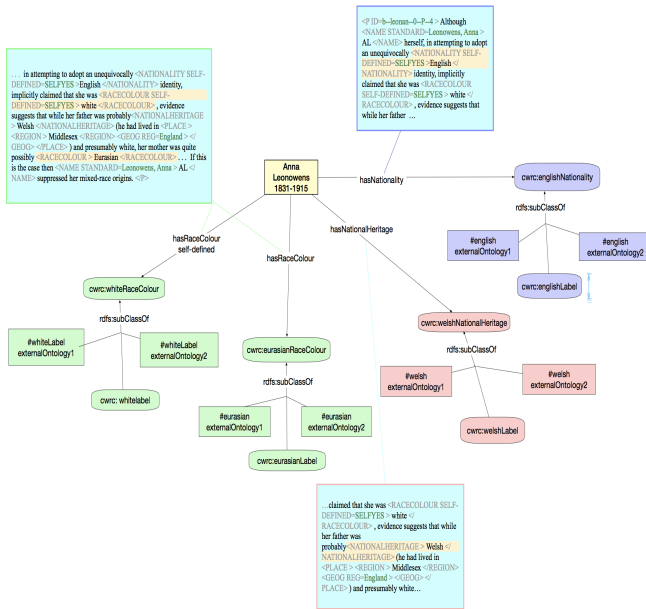


Figure 4: Cultural Formation triples related to Anna Leonowens, with corresponding XML-encoded context notes

Figure 5 indicates the ability to see patterns and outliers related to different categorizations of Jewishness in a small subset of Orlando authors.

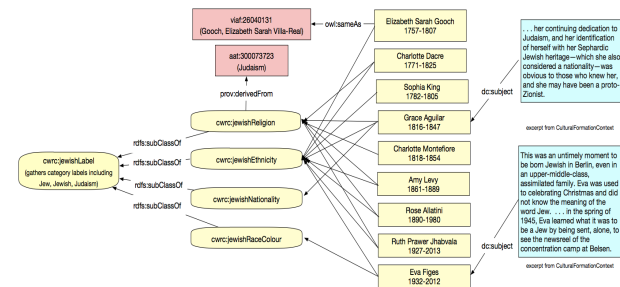


Figure 5: Subset of CF triples related to a subset of writers, with sample context annotations and external links; predicates linking individuals to subclasses are inferred (e.g. the edge between Elizabeth Sarah Gooch and cwc:jewishReligion is hasReligion)

Our live presentation will demonstrate the ontology in action using the interactive HuViz (Humanities Visualizer) interface with a larger dataset.

## Ontologies

- CWRC ontology: <http://sparql.cwrc.ca/ontology/cwrc>
- CWRC sparql end point: <http://sparql.cwrc.ca/>
- Orlando Biography schema containing Cultural Formation tagset: <https://github.com/cwrc/CWRC->

- Fuss, D.** (2013). *Essentially Speaking: Feminism, Nature & Difference*. New York: Routledge.
- Gold, M.** (ed.) (2012). *Debates in the Digital Humanities*. Minnesota: University of Minnesota Press.
- Gold, M. K., and Klein, L. F.** (eds.) (2016). *Debates in the Digital Humanities 2016*. Minnesota: University of Minnesota Press.
- Johnston, J.** (1973). *Lesbian Nation: The Feminist Solution*. New York: Simon and Schuster.
- Lerman, J.** (2013). "Big data and its exclusions." 66 *Stanford Law Review Online* 55: 55-63. [http://www.heinonline.org/subzero.lib.uoguelph.ca/HOL/Page?handle=hein.journals/slro66&start\\_page=55&collection=journals&id=66](http://www.heinonline.org/subzero.lib.uoguelph.ca/HOL/Page?handle=hein.journals/slro66&start_page=55&collection=journals&id=66) (accessed April 7, 2017).
- McPherson, T.** (2012). "Why are the Digital Humanities so white? Or thinking the histories of race and computation." In M. Gold (ed). *Debates in the Digital Humanities*. Minnesota: University of Minnesota Press, pp. 139-160.
- Muninn Project.** "Appearances Ontology Specification - 0.1." 2012. <http://rdf.muninn-project.org/ontologies/appearances.html>
- Munt, S.** (1998). "Sisters in exile: the lesbian nation." *New Frontiers of Space, Bodies and Gender*. London: Routledge, pp. 3-19.
- Nakamura, L.** (2002). *Cybertypes: Race, Ethnicity, and Identity on the Internet*. London: Routledge.
- Ross, B.** (1995). *The House that Jill Built: A Lesbian Nation in Formation*. Toronto: University of Toronto Press.
- Smith, J.** (2016). "Working with the Semantic Web." In C. Crompton, R. J. Lane, and R. Siemens (ed.). *Doing Digital Humanities: Practice, training, research*. London: Routledge, pp. 273-88.
- Treviranus, J.** (2014). "The value of the statistically insignificant." *Educause* 49:1. <http://er.educause.edu/articles/2014/1/the-value-of-the-statistically-insignificant> (accessed 7 April 2017).
- W3C.** (2017). Web Annotation Data Model. 23 February 2017. <https://www.w3.org/TR/annotation-model/> (accessed: 7 April 2017).
- Wikipedia contributors** (2017). "Ontology (information science)," *Wikipedia, The Free Encyclopedia*. [https://en.wikipedia.org/w/index.php?title=Ontology\\_\(information\\_science\)&oldid=772391479](https://en.wikipedia.org/w/index.php?title=Ontology_(information_science)&oldid=772391479) (accessed April 7, 2017).
- Woolf, V.** (1938). *Three Guineas*. London: Hogarth Press.