

# The Curious Identity of Michael Field and its Implications for Humanities Research with the Semantic Web

Susan Brown

*University of Alberta / University of Guelph  
Implementing the New Knowledge Environment/  
Text Mining & Visualization for Literary History  
Edmonton, Canada  
susan.brown@uoguelph.ca*

John Simpson

*University of Alberta  
Implementing the New Knowledge Environment/  
Text Mining & Visualization for Literary History  
Edmonton, Canada  
john.simpson@ualberta.ca*

**Abstract**—This paper uses the case of author Michael Field, the shared writing identity of two late Victorian women, to consider the implications of embracing the semantic web for humanities research. It is argued that the ontologies prevalent today reveal a lack of nuance when it comes to the complex relationships that are the focus of much humanities research, such as the connection of names to persons, particularly with respect to authorship. Further, the current state of ontology use aside, even the sophisticated use of OWL, SKOS, or ontology alignment techniques for linking big semantic web collections stands to hinder humanities research by hiding rather than exposing difference. We use the outlier Michael Field to highlight what much of the valuable work of the humanities is about and in doing so bring to the fore the challenge of formalizing complex social meanings that can otherwise be overlooked or dismissed as a trivial technicality. As a solution the humanities community is encouraged to begin engaging more directly in the construction of semantic web tools and infrastructure.

**Keywords**—ontologies; semantic web; humanities; Michael Field; pseudonyms

## I. HUMANITIES SCHOLARSHIP AND THE SEMANTIC WEB

Humanities scholars have long recognized the potential of linking datasets to allow illumination of subject areas via computer-based analysis and for providing new forms of access to the vast array of texts and other materials where much of our intellectual and cultural history resides. While the majority of primary and secondary material collections in digital form remain effectively siloed, immense strides have been made towards freeing these over the past few years, enabling us to begin investigating “the kinds of humanistic phenomena” that “appear only at scale” [28].

The path to the current state of affairs began with coining “semantic web” in 2001 [3] and since then the number of data sources available in formats suitable for linked open data has increased exponentially. In 2007 there were just 12 data sets published as linked open data, but by 2011 this number had increased to 295 [27]. As of writing (August 2013) the CKAN data hub currently boasts 6,136 data sets meant to be connected and shared [11]. After a period of caution, we are now witnessing a decided increase in the

uptake of linked open data standards and semantic web technologies for humanities research. This uptake is expressed in three sorts of contributions: aggregations of humanities relevant information; specialty topic explorations; and tools [7].

Aggregator projects collect together massive amounts of information providing a one-stop-shop of sorts for information attached to a particular theme or topic. They archive a wide range of historical materials for aggregation and discovery, such as Europeana or PELAGIOS, or provide a more focused but no less extensive collection of materials by type or theme, such as is done by NINES.

There also exist much smaller projects that stand out because of the special attention that is given to curating the information that they hold. What they lack in breadth they attempt to make up for with depth and/or prototyping alternative ways to interact with the information they present. Linked Jazz and InPhOare exemplars of this sort of humanities contribution to linked data and the semantic web.

More generalized tools are also being built to allow humanists to contribute to interact with and contribute to semantic web materials, including Pundit, SharedCanvas, AustESE, and CWRC-Writer. These tools have all been developed in relation to humanities or memory institution projects.

These sorts of projects unify like-minded communities, reflecting the premises of the semantic web:

[T]he Semantic Web is built up from small like-minded communities that can find agreement on terms amongst themselves. Applications, then, can and do interact without attempting to achieve global consensus. There is no requirement for global ontologies: instead, an application need only map the terms relevant for a particular transaction into a common vocabulary. Of course, though agreement need only be local, adoption of existing vocabularies facilitates data sharing and integration. [52]

What humanists ultimately want from the semantic web is not only access to all material, and only that material, of interest to a particular inquiry but also the ability to extract from the massive aggregation of separate datasets new leads, connections or insights. That is, they subscribe to the hope that it will be possible to use the semantic content and structures of semantic web data to produce knowledge that could not be arrived at by other means. To address these demands requires a balancing act between collecting as much as we can and ignoring what is not of interest, between errors of admission and errors of omission, and between taking too much and taking too little. Currently we are erring on the side of taking too much as we gather together masses of information in the twin hopes that we can sort out what is of value later and that the aggregation itself will produce information of value. Such are the drives of the big data movement.

Within this context, we consider how the development of the semantic web is shaping up for the humanities. Much of what is offered here emerges from work being done to connect a significant humanities datastore with the semantic web. The datastore—published as *Orlando: Women's Writing in the British Isles from the Beginnings to the Present*—is a born-digital reference and research resource comprised of more than 1,300 detailed biocritical entries, capturing over 27,000 individual people plus extensive historical and bibliographical detail [41], [8]. The prose within each entry has been marked up using XML to identify information related to authors' lives and writing careers, contextual material, timelines, sets of internal links, and bibliographies [6]. With the integration of the Orlando Project data with the semantic web as the background for this inquiry, we first summarize one major shortcoming of current approaches to building humanist data stores for connectivity to either the linked data cloud or the semantic web. After briefly outlining the particular importance of naming and the representation of identities, we then provide the specific example of the late Victorian writer Michael Field, exploring the challenges it poses to the treatment of personal identity within a semantic web context. We argue that even the most sophisticated use of ontologies like OWL or SKOS, or the use of ontology alignment techniques, will not allow linked data and the semantic web to meet some fundamental requirements of humanities researchers. The current approach to the semantic web, with its background assumptions and practical implementations, actually runs counter to the achievement of our long-term goals. This problem can be solved at least in part, we argue, if humanities researchers revise both our expectations and the direction of the development of data repositories, inference engines, and ontology standards.

## II. THE EROSION OF DIFFERENCE THROUGH AGGREGATION

Granting that linked (open) data and the semantic web are both still in their relative infancy or just past what the Gartner Group refers to as the peak of inflated expectations [21], [37], there are still elements about the current state of affairs and its future tendencies that are of concern to humanists. Indeed, both the potential and the shortcomings of existing large data sets for much humanities research have received much attention, particularly as manifest in the concepts of “distant reading” and “cultural economics” [35], [36], [38]. While there are many challenges to integrating humanities data with the semantic web, the focus here is on the erosion of markers of difference through aggregation.

Humanists research diverse aspects of human culture and cultural processes on both large and small scales, with a particular emphasis on the relationship of cultural processes and products to their connections with human agents. To the extent that RDF is built to work with and through machine-processable ontologies, the implications for knowledge representation and knowledge construction are massive, and are a major source of the attraction to linked data frameworks. For the humanities the move to linked open data and the semantic web is potentially paradigm-shifting, presenting a tantalizing prospect of the opportunity to access, investigate, and make sense of human culture on an unprecedented scale.

The flexibility of linked data technology lies in the fact that each datastore can be given its own vocabulary and ontology to suit its needs and link out to other datastores as desired, at least *in principle*. However, linking up datastores often means connecting different vocabularies and ontologies. This in turn brings with it a pressure towards generalization rather than specificity because of the deep complexities involved in both generating and aligning ontologies of any non-trivial sophistication. While there are sophisticated ontologies being developed and implemented for specific domains and projects, the norm by far for most semantic web projects, including humanities ones, is to cherry-pick terms from existing ontologies that strike them as generally useful for cobbling together their own vocabulary [29], [46]. In the actual implementations of most semantic web datastores and aggregators, precision and richness are often foregone in favour of matching up with widely-used ontologies. This translates into generalization and loss of specificity, making the aggregated data less useful for humanities inquiry than it might otherwise be. Dominic Oldman of the British Museum eloquently summarizes the situation:

[M]any linked data points lack the quality to be more than basic information jukeboxes that, in turn, support applications with limited usefulness and shelf life. In short, the current cultural heritage linked data movement, concentrating on access (a fundamental objective), may have ignored some of

reasons for establishing networks of knowledge in the first place. [39]

The rationale for linking data, he implies, includes not merely ease of access, the low-hanging fruit enabled by the consolidation of datasets, but the larger prize of seeing culture otherwise than we can with our current data and methods. It is certainly true that “a little semantics goes a long way,” [23] but to have more than catalogues that erase the specifics of difference in the interests of aggregation we need a little bit more than “a little semantics.”

Difference, as in distinctions of particularity, specificity, and locality of meaning, provides us with the ability to make sense of the complexities of human records and human lives. Difference also means difference from the norm: the marginalization and anomalousness associated with categories such as race, gender, or nation are registers of difference. Difference is key to understanding how literature, culture, and ideas develop. The semantic web, as a tool for connecting and representing human knowledge and understanding on a massive scale, is poised to either promote or hide difference, and differences in approach matter.

Consider a couple of the most widely used ontologies. FOAF, the Friend of a Friend ontology that was originally intended as a supplement/alternative to social networking websites, describes what constitutes a “person” thus:

The Person class represents people. Something is a Person if it is a person. We don't nitpick about whether they're alive, dead, real, or imaginary. The Person class is a sub-class of the Agent class, since all people are considered 'agents' in FOAF. [5]

This “no nitpicking” approach ensures that the ontology is flexible enough to appeal to many users by covering all possible understandings of personhood. A similar move towards broad generalization is made in the Dublin Core Metadata Initiative [15], an ontology directed towards capturing the relationships that persons have to the things that they create. Here there are only two roles which can be assigned to a person: they must either be a creator or a contributor. DCMI does this to maximize its applicability to a wide range of works and to simplify the application of these properties to the persons involved. There is no need to decide if a person is an artist or an author or an architect: they are all brought together under one or the other of these two terms.

Clearly, neither of these two highly generalized approaches reflects the kind of differences routinely recognized by humanists in the course of scholarly inquiry. If we presume that such distinctions are fundamental to humanist work, it is clear that the mountains of data we are amassing using such ontologies will support only very limited types of inquiry or inference.

### III. WHAT'S IN A NAME?

Names are important because the act of naming allows us to differentiate the particular from the general in our

thoughts and ideas. Names thus produce affordances with which we are able to separate, capture, and manipulate tokens from a sea of types. As Hope Olson argues in her critical study of library subject representation: “Naming is the act of bestowing a name, of labelling, of creating an identity” [40]. Indeed, much of the power of the semantic web comes from the ability to name entities and to link materials across disparate sources.

Olson's observation of naming as an act of identity creation suggests an important bridge between naming and personhood through the concept of identity. In many instances naming is a foundational act of creating persons specifically because it provides an identity where there previously was none, a simultaneous creation of an individual and an extension of the group bestowing the name. At times this act of naming leads the way to personifications that challenge our understanding of personhood in complex ways (consider how the naming of both corporations and stuffed animals leads us to lean towards treating them both as persons in potentially controversial ways). Our concern here is that personhood is both a complex and a crucial characteristic that ontologies must be designed to capture appropriately, lest some of our understandings of the world fail to be addressed.

What is at stake in such designs is the ability of the resulting ontology to capture adequately not just the nuances of a relationship between persons, organizations, objects, and other such things, but for these relationships to be captured in ways that are even broadly correct. As a trivial example, consider an ontology that merged FOAF and the DCMI (a reasonable assumption since these are two of the most popular ontologies [29]). In doing so, an ontology designer would be left with only two roles in the DCMI to equate to the people captured in FOAF: they could be either a `dc:contributor` or `dc:creator`. Regardless of how this combination is pitched, it amounts to a flattening out of relationships that makes detailed inquiry into authorship and reception an extremely difficult prospect. This flattening is not so much a matter of mistaken ontology combination as it is poor ontology selection. For poor ontology combination in action, let us consider the case of Michael Field.

### IV. MICHAEL FIELD

Humanists have been at work for more than two decades, since the days of the earliest graphical browsers, to populate the web with more diverse content about women writers. A major aim of the Orlando Project, for instance, was specifically to increase the representation of women both on the web and in literary history [8]. This concern goes beyond academia: the *Rewriting Wikipedia Project* initiated in May 2013 tackled the well-known gender gap in the most commonly used encyclopedia in the world, particularly with respect to women from marginalized groups [25], [20]. So there remains a strong impetus to attempt to rectify the

omissions and biases of history and of mainstream culture by promoting knowledge of women’s achievements.

Women also constitute a particularly good class of people through which to consider relationships between names and identities, given that the institutionalization of women’s name changes through marriage in many Western countries means that an individual woman often uses several names over the course of a lifetime. This makes it hard to connect all the different names to the same historical individual, presenting a long-standing problem with respect to library cataloguing, census research, and other indexing systems. Women writers present an even better test case for testing the ability of semantic web representations to handle complex relationships between names and identities. Virginia Woolf remarked in 1928, “I would venture to guess that Anon, who wrote so many poems without signing them, was often a woman” [50], [45], and considered it an indication of the extent to which writing has been fraught for women in Western European culture. Much inquiry into the gendering of the field of literary production and reception has focused on the extent of, the reasons underlying, and the impact of the adoption of anonymity and pseudonymity by female authors [18], [34], [16], [30].

The author Michael Field was the product of an artistic collaboration between Katherine Harris Bradley and Edith Emma Cooper during the late nineteenth and early twentieth centuries. Bradley and Cooper are marginalized and outliers in several ways. An aunt and niece by marriage, they were avowed lovers in a historical period in which lesbianism, even less than homosexuality, was without a public name or identity. They represent an unusual type of authorship, having written together under the name and the shared identity of “Michael Field” after first publishing under the dual pseudonyms of Arran and Isla Leigh [8]. These two biological and historical entities operated as a single author, but they carried their shared pseudonym also over into their private life, in which Katherine was Michael and Edith variously Field, Henry, Hennery, Henny, or Hennie-boy [13]. Then there is their relationship to the name of Sappho, the ancient Greek poet at the head of the lyric tradition of Western poetry whose name was becoming a rallying point for an emergent lesbian subculture [44]. Sappho’s remaining fragmentary texts (a potent example of the extent to which humanities data can never be complete) were literally incorporated into the lyric poems of the Michael Field’s *Long Ago* in such a way that, as Yopie Prins argues, “Bradley and Cooper appropriate Sappho as a name simultaneously proper and improper, their own and not their own” [43].

Field’s name is thus layered and complex in its relationship to historical personhood: in addition to Bradley and Cooper, there is also Sappho, about whom so little is known that she is almost as much of a construction as Michael Field, and the familiar variants of the pseudonym used by the two women in everyday life. The properties of Field

are equally complex in relation to both gender and number. Holly Laird [26] uses a female pronoun, while Lisa Prins [43] and *Orlando* [41] use the third-person plural, and yet the name was clearly chosen by the women as a singular, masculine authorial identity invoked as “him”. Male pseudonyms adopted by women are interestingly bisexual, in that they carry in their motivations traces of the socialized femininity of the author even as they publicly perform masculinity. The singularity of the pseudonym, however, might equally also be applied to the female union of Bradley and Cooper, who considered themselves married [19]. As Lorraine York notes, this “pronominally irregular relationship” led Robert Browning to send, via a mutual friend of Field, greetings to “both of him” [51]. A unifying view is manifestly wrong. So too is a divisional view. And like much of humanities research, saying that the truth is somewhere in between is of no help at all.

As Laird comments, “Field in fact anticipates the feminist, historicizing scholar, the scholar who seeks representations of women and gender in the fracturing mirror of past texts in order to put the fragments together in her own documents” [26]. In other words, to smooth out and create a unified identity from the differences embedded in the signature of Michael Field would be to miss the point of that signature and erase history. Let us turn then to the view of Michael Field that we gain from existing linked data.

DBpedia is a Linked Data project based on structured information derived from Wikipedia entries [2], [12]. Looking at the DBpedia entry for Michael Field, visualized as network graph in Figure 1, begins to reveal the sorts of problems that can arise in using RDF to describe such a complex entity.

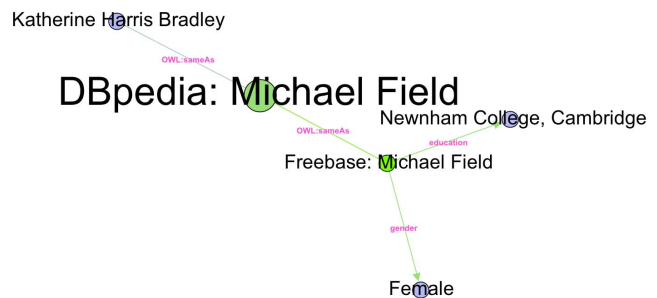


Figure 1. DBpedia sends Michael Field to College

In the representation provided in DBpedia, Michael Field takes on interesting properties based both on what is said and on what is not said. To begin, Field is asserted to be identical with Bradley. This makes it the case that any property that is attributed to her gets attributed to Field because OWL defines the sameAs relationship as satisfying the strongest identity claim possible [52]. DBpedia goes further, reaching outside itself and linking to Freebase, another RDF

based knowledge store. Freebase asserts that Field went to Newnham College and is female. Of course neither of these is true. It was Bradley who went to Newnham—a full ten years before her collaboration with Cooper produced Field—and Field is not female (at least not female with the certainty asserted here by DBpedia/Freebase, given the complexities outlined above).

Further, based on the DBpedia representation a reasonable take-away would be that Field is the actual person and perhaps Bradley is the artistic creation. But note that not even inverting these assertions makes things right because the DBpedia approach does not include the contributions of Cooper at all.

Even a scholarly literary resource dedicated to the history of difference has trouble capturing Field. The scholars behind *Orlando* privileged the publishing persona that arose from the collaboration between Bradley and Cooper, and therefore produced a lengthy entry on “Michael Field”, rather than entries on each woman individually. This entry is marked up with XML using the project’s bespoke schema. However, while the prose captures the complex nature of Field as an entity, the XML does not. As translated into RDF from the *Orlando* schema, Field ends up being born twice and dying twice, and is made out to be female, when the intention behind the partnership was to allow the women to operate as male within the literary scene of the day. So the RDF extracted from the entry is both factually inaccurate and impoverished. The network graph captured in Figure 2 should make the limitations clear.

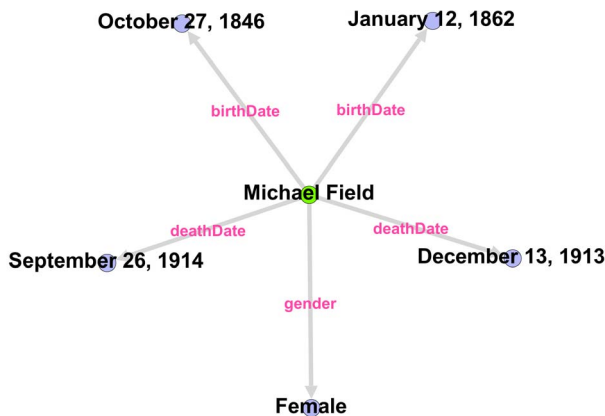


Figure 2. *Orlando* allows that Michael Field was born and died twice

## V. IMPLICATIONS FOR THE HUMANITIES AND THE SEMANTIC WEB

At this point it may seem that the problem here is that the wrong schemas and ontologies are being used or that there is no *real* issue here because Field is such an outlier that invoking him—or her—or them—amounts to scholarly nitpicking. Michael Field might also be seen as

the sort of problem that can be solved by some alternative use of the tools at hand (e.g. not using owl:sameAs) or some other simple/clever technical solution. We urge caution against comfortably settling into any of these dismissive conclusions.

Not every problem that has arisen from this process is amenable to a technical solution. The author Michael Field brings to the fore the extent to which the construction of ontologies poses a real challenge. While Field could be shoehorned into existing ontologies as either a pseudonym or a corporate author, we contend that neither is an adequate representation of the relationship, even in combination. Indeed Field has been pretty much shoehorned into the DBpedia ontology with the results outlined above. Though it is in some respects highly unusual, in other ways Field’s case is far from unique. Fuzziness with respect to the relationship between names and cultural creators entangles authorship and personhood in ways that result in confusion and outright error in numerous semantic web representations.

The case of Field is not a mistake that can be easily cleared up: it is not as if someone has mislabeled a dog fish as a dog. Rather, the problems arise because the DBpedia, Freebase, and Orlando Project ontologies apply properties related to personhood in commonsensical and normalizing fashions that are inadequate for capturing the complexities of Field. Field is, in sum, an entity who challenges the very concept of personhood. We get by with our everyday understandings, in life and in most Linked Data contexts, because most cases do not give rise to the kinds of strangeness we see in the RDF representations of Field. We could take the FOAF “no nitpicking” route and just let everything in, but that avoids the problem. Field is the ontological equivalent of approaching the speed of light, a fringe case beyond everyday experience, indicating that our common sense understandings and knowledge may not be as perfect as we treat them as being, both within a semantic web context and elsewhere.

An adequate representation of the particular kind of personhood and authorship embedded in the signature of “Michael Field” seems to us impossible within the terms of existing bibliographical ontologies or other more general ontologies. Nor is this a case of sloppy popular ontologies vs. rigorous expert ontologies: neither the widely employed FRBR (Functional Requirements for Bibliographic Records) nor PRO (Publishing Roles Ontology), is a help in tackling this problem. Even constructing a new ontology to account for the particular challenges posed by Field would only serve to address the symptom rather than the underlying cause: the messiness hidden behind our everyday ontological assumptions.

Any nuance that might be built into a new ontology, moreover, would likely be obscured if that data were brought into use alongside other data sets, since in that context it would likely be generalized to match up with the limited

vocabulary of a more standard ontology such as Dublin Core. An example of this kind of effect is that in the Virtual International Authority File (VIAF), a major resource comprised of the records of leading national libraries worldwide, twenty-first-century editor Sharon Bickle shows up as co-author with Michael Field, the Victorian author she edited [48]. Sappho and composer Johann Strauss, Jr are also listed as co-Authors. This is not an insignificant problem on a small scale and if Swoogle is to be believed and there are really over 10,000 namespaces as part of the semantic web [47] the problem is anything but small.<sup>1</sup>

Again, it may be tempting to assert that this is merely a technical problem that can be solved by the appropriate use of an ontology like SKOS, the clever use of OWL, or even the creation of a new ontology custom built for the purpose of making such problems go away. What such solutions really amount to are acts of ontology integration and while technically plausible in theory in practice they present significant challenges that only well funded projects have the capacity to pursue.

Ontology integration is a well-known and growing problem on the semantic web for which there are no general solutions that come without costs. The most robust of these solutions is what Pinto et al. count as integration proper [42]. This approach assembles a new purpose-built ontology from the pieces of other ontologies. The connective logic is worked out at each step and a new ontology results. Understandably, there is a great deal of work and reflection that needs to go into this process, and so this approach is not undertaken lightly if accuracy is even remotely important. A second approach discussed by Pinto et al. is merging. Here existing ontologies are combined together in a way that produces a new ontology but without dismantling the original ontologies. OWL offers two ways to do this: use *owl:imports* or use *owl:sameAs*. Again, both come at a price.

*owl:imports* amounts to taking the union of all the ontologies that fall under its scope. Of course, this opens the door to serious problems, not the least of which is that it fails to allow for filtering of any sort, simply accepting other ontologies—and any uses they have made of *owl:imports*—outright, paving the way for ontology explosions [22]. *owl:sameAs* allows a property from one ontology to be equated with another in the strictest sense. Using this powerful declaration brings with it all the problems of the *owl:import* but with the additional inheritance of the features of each property into the other.

There are other approaches but they all amount to the same thing, finding new ways to relate terms used within one context to the terms used in another. What is important to note here is not that potential solutions already exist in the database and knowledge base communities, but that the

humanities community is not engaging with these solutions on any significant scale. Rather, the humanities community is focused on building large repositories and continues to choose the assert the most threadbare vocabularies as the means of enabling this goal. As suggested earlier, the consequence of this will be information juke boxes, repositories that are able to dish up preselected information through generic interfaces, but little more. If we are really interested in making the semantic web a tool which will allow us to see interesting things that we could not perceive otherwise, then we need to recognize that limiting the complexity of vocabularies and ontologies so that we can easily collect as much information as possible together will not achieve this end.

What the humanities community needs to be doing on a larger scale is the following:

- 1) Constructing more custom ontologies directed at capturing humanities-relevant information rather than adopting those built by other communities for other purposes;
- 2) Pursuing the ontology integration techniques already developed to address the challenges of integration, despite their complexity and the real possibility that they will not work;
- 3) Designing inference engines specifically for humanities research.

What these activities amount to is engaging in the practice of building and contributing to the infrastructure of the semantic web rather than the practice of adding new repositories to an already existing infrastructure. The erosion of difference as a result of the tendency towards homogenization that is evident in large data aggregations is not the only challenge that humanists will need to overcome in order to work meaningfully with the semantic web. Others include overcoming the assumed facticity and completeness of the information that is being aggregated [14], and the limitations of the deductive logic framework that is assumed by the majority of ontologies and inference tools available [32]. If we do not engage with the complexity of the challenges facing us with respect to data aggregation, then we should not be surprised when the joy of being able to do faceted searches in a one portal rather than many wears off, leaving us wondering what happened to the promise of the semantic web to allow us to discover that which we did not already know, to help us plumb the specificities of our vast stores of knowledge for new insights.

## VI. CONCLUSION

In sum, the semantic web ontologies prevalent today reveal a lack of nuance when it comes to the complex relationships between names and persons, particularly with respect to authorship. Moreover, the generalizations required to link semantic web collections risk weakening the ability of their data representations to reflect the complexities of

<sup>1</sup>Even if the more conservative estimates of others [46], [10], [29] are more accurate the challenge of integration remains significant.

cultural artifacts and processes. Tim Berners-Lee stresses the extent to which linking across specific vocabularies, in his view, takes precedence over specific vocabularies:

The library community has historically seen the Semantic Web as primarily about metadata. While that is important, it is only one aspect of the larger picture. There is financial data, chemical data, biotechnology data, experimental data, geographic data and more. All of these domains have their own vocabularies, with few explicit points of connection. The Semantic Web is aimed at bridging those gaps, and allowing links across fields. Libraries have long understood the importance of established vocabularies, and have led in their development. To the extent that data can be encoded in common syntaxes like RDF and described with public vocabularies, they can be more accessible and more useful. People and applications can draw better correlations, better connections, better inferencing, and these can lead us to more effective use of information. [49]

“Public” here really equates to a standardized vocabulary, or linked to a standard vocabulary, which in effect means that local vocabularies need to be equated or generalized to standard vocabularies in order for the promised amelioration to take place. RDF is clearly emerging as a standard that is effecting significant changes in how data is represented, exchanged, and used on the web, and it relies on standard ontologies to serve as the bridges between specialist vocabularies. Standard ontologies can thus enable technological progress, but those devising and employing them need also to be mindful of the negative impacts of standardization powerfully articulated by Geoffrey Bowker and Susan Leigh Star:

Information scientists work every day on the design, delegation, and choice of classification systems and standards, yet few see them as artifacts embodying moral and aesthetic choices that in turn craft people’s identities, aspirations and dignities. [4]

While many standards, particularly technical ones and ones associated with the natural sciences, have arisen without the kind of attentiveness to difference outlined above, the critique of such standards is growing: witness the recent success of Melissa Terras in having the *Text Encoding Initiative* guidelines, which previously embedded the encoding of women literally as the second sex, revised [1]. Other related initiatives consider how best to model fluid difference, along with explorations of how encoding practices can leverage established standards without acceding to their terms [31]. Such work intersects with the Orlando Project’s attempt to translate into RDF its tagset for encoding “cultural formation,” which is attentive to historical and cultural specificity

related to such identity categories as sexuality, ethnicity, race and colour, and linguistic roots [9]. More theoretically, Tara McPherson has begun investigating the ideological implications of the base technologies, such as the UNIX operating system, that are shaping our digital work in the humanities [33].

The questions raised here are important given the increasing uptake of semantic web technologies by research projects in the humanities, and by resource providers such as libraries and museums. As part of this process of adoption, many existing ontologies initially developed for other technologies and purposes are being translated into semantic web form to enable the leveraging of existing metadata in a semantic web context. XML schemas in particular are susceptible to such translation, and there is increasing production of RDF based on existing XML markup. Indeed the W3C offers the “ConverterToRdf” tool meant to facilitate such translations, whether the migration is to RDF or to the ongoing exposure of XML data in RDF form. Moreover, resources are increasingly being interlinked not by human beings but by algorithms. For instance, in late 2012 a VIAF bot successfully added 250,000 links to Wikipedia [48]. The result will be a much more densely interlinked and, one hopes, interoperable semantic web than the current one, but its structure will be increasingly determined by its own vocabularies and ontologies, which will in turn increasingly determine what can be represented meaningfully on the web. If this sounds circular, that’s because it is.

The case of Field helps to lay bare the assumptions about the nature of authorship in relation to personhood that operate quite naturally, as culturally-specific ideological systems do, within our understanding of literary relations, and the challenges that the humanities’ interest in the unique, the outlier, and the limit cases of culture pose to a semantic web framework. This preliminary examination of the limitations of existing ontologies, both the one that we are deriving from *Orlando*’s bespoke XML markup and well-established ones like Dublin Core, suggests that they will not adequately handle border cases, despite the fact that marginality and uniqueness are what humanities scholars often seek to discover and analyse.

We are still in the initial stages of investigating the implications of semantic web representations for humanist research, particularly as concerns the ability of such representations to reflect humanist epistemologies, or ways of knowing, adequately. The representation of basic names of persons within the semantic web, of their relationships to other names, to properties, and to creation, publishing reception, or authorship, have implications that reach far beyond literary studies or book history to the heart of our information systems. The problem does not have to do so much with RDF as a form of data representation, but with how it is being deployed. Challenging as these cases are, the flexibility of the Resource Description Framework that



is leading to these confusions is also paradoxically our best hope of arriving at strategies for representing difference effectively across large sets of data. Outliers such as Michael Field bring to the fore the challenge of formalizing complex social meanings such that the resulting web of information can operate with a difference.

#### ACKNOWLEDGMENT

This paper emerges from the work of the Implementing New Knowledge Environments (INKE) project and the Text Mining and Visualization Project, both sponsored by the Social Science and Humanities Research Council of Canada. We are grateful to our INKE colleagues (Harvey Quamen, Jon Bath, Lisa Goddard, Jon Saklofske, Jentery Sayers, Adèle Barclay, Alex Christie, and Mandy Elliott) for collaborations that inform this paper, and to Emily Earhart and Laura Mandell for exchanges on related topics.

#### REFERENCES

- [1] "#425 TEI using outmoded ISO 5218 for sex value attribute." Sourceforge discussion. 21 Jan. 2013-14 April 2013. Web. 10 May 2013. <http://sourceforge.net/p/tei/feature-requests/425/>.
- [2] Auer, Sren, et al. "Dbpedia: A nucleus for a web of open data." *The semantic web*. Springer Berlin Heidelberg, 2007. 722-735.
- [3] Berners-Lee, Tim et al. "The semantic web." *Scientific American* 284.5 (2001): 34-43. Web. 10 May 2013. <http://www.scientificamerican.com/article.cfm?id=the-semantic-web>.
- [4] Bowker, Geoffrey, and Susan Leigh Star. *Sorting Things Out: Classification and its Consequences*. Cambridge: MIT Press, 1999. Print.
- [5] Brickley, Dan, and Libby Miller. "Person." FOAF Vocabulary Specification 0.98. FOAF Vocabulary, 2010. Web. 10 May 2013. [http://xmlns.com/foaf/spec/#term\\_Person](http://xmlns.com/foaf/spec/#term_Person).
- [6] S. Brown, S. Fisher, P. Clements, K. Binhammer, T. Butler, K. Carter, I. Grundy, and S. Hockey, "SGML and the Orlando Project: descriptive markup for an electronic history of women's writing," *Computers and the Humanities*, vol. 31, no. 4, pp. 271-85, 1997.
- [7] Brown, S., H. Quamen, J. Simpson, J. Bath, L. Goddard, J. Saklofske, J. Sayers, A. Barclay, A. Christie, M. Elliott, and the INKE Research Group (November 2012). *E/Merging Models for the Production of Online Research through Linked Data*. Research Foundations for Understanding Books and Reading in the Digital Age: E/Merging Reading, Writing, and Research Practices, Havana, Cuba.
- [8] Brown, Susan, Patricia Clements and Isobel Grundy, ed. *Orlando: Women's Writing in the British Isles from the Beginnings to the Present*. Cambridge: Cambridge University Press, 2010. Web. 10 May 2013. <http://orlando.cambridge.org>.
- [9] Brown, Susan, Patricia Clements, and Isobel Grundy. "Sorting Things In: feminist knowledge representation and changing modes of scholarly production," *Womens Studies International Forum* 29 (2006): 317-25
- [10] Cheng, Gong, Min Liu, and Yuzhong Qu. *NJVR: The NanJing Vocabulary Repository*. In *Semantic Web and Web Science*, 265272. Springer, 2013.
- [11] Data Hub. CKAN, 2012. Web. 10 May 2013. <http://datahub.io/>.
- [12] "DBpedia." <http://dbpedia.org/About>.
- [13] Donoghue, Emma. *We Are Michael Field*. Absolute Press, 1998.
- [14] Drucker, Johanna. "Humanities approaches to graphical display." *Digital Humanities Quarterly* 5.1 (2011). Web. 10 May 2013. [http://www.johannadrucker.com/pdf/hum\\_app.pdf](http://www.johannadrucker.com/pdf/hum_app.pdf).
- [15] "Dublin Core Metadata Initiative." DCMI, 2013. Web. 10 May 2013. <http://dublincore.org/documents/2012/06/14/dcmi-terms/?v=terms>.
- [16] Easley, Alexis. *First Person Anonymous: Women Writers and Victorian Print Media, 1830-1870*. Ashgate Publishing, Ltd., 2004. Print.
- [17] Europeana. Web. 10 May 2013. <http://www.europeana.eu/>.
- [18] Ezell, Margaret JM. *Writing women's literary history*. Johns Hopkins University Press, 1996. Print.
- [19] Field, Michael. *Works and Days*. Ed. Thomas Sturge Moore, and D.C. Sturge Moore. J. Murray, 1933.
- [20] Filipacchi, Amanda. "Wikipedias Sexism Toward Female Novelists." *The New York Times Sunday Review*. *The New York Times*, 24 April 2013. Web. 10 May 2013. [http://www.nytimes.com/2013/04/28/opinion/sunday/wikipedias-sexism-toward-female-novelists.html?smid=tw-share&\\_r=0](http://www.nytimes.com/2013/04/28/opinion/sunday/wikipedias-sexism-toward-female-novelists.html?smid=tw-share&_r=0).
- [21] Gartner Group. "Hype Cycles." Web. 6 August 2013. <http://www.gartner.com/technology/research/methodologies/hype-cycle.jsp>.
- [22] Grau, Bernardo Cuenca et al. *Working with Multiple Ontologies on the Semantic Web*. SWC 2004, LNCS 3298, pp. 620634, 2004.
- [23] Hendler, James. "A Little Semantics Goes a Long Way." Web. 6 August 2013. <http://www.cs.rpi.edu/hendler/LittleSemanticsWeb.html>
- [24] Klein, Max, and Melissa Renspie. "VIAFbot Edits 250,000 Wikipedia Articles to Reciprocate All Links from VIAF into Wikipedia." OCLC Research, 2012. OCLC. Web. 10 May 2013. <http://www.oclc.org/research/news/2012/12-07a.html>.
- [25] Koh, Adelina, and Roopika Risam. "The Global Women Wikipedia Write-in #GWWI." *Postcolonial Digital Humanities blog*. Web. 10 May 2013. <http://dhpoco.org/2013/03/21/the-global-women-wikipedia-write-in-gwwi/>.
- [26] Laird, Holly. "Contradictory legacies: Michael Field and feminist restoration." *Victorian poetry* 33.1 (1995): 111-128. Web. 10 May 2013.



- [27] "The Linking Open Data cloud diagram." LOD, 2011. Web. 10 May 2013. <http://lod-cloud.net/>.
- [28] Liu, Alan. "The State of the Digital Humanities: A Report and a Critique." *Arts and Humanities in Higher Education* 11.1 (2012): 1-34. Web. 10 May 2013. <http://ahh.sagepub.com/content/early/2011/11/30/1474022211427364.full.pdf+html>.
- [29] LODstats, <http://stats.lod2.eu/stats>
- [30] Mandell, Laura. "Feminist Critique vs. Feminist Production in Digital Humanities," Invited, Keynote Address, "Womens History in the Digital World," 22 March 2013, Bryn Mawr University, The Albert M. Greenfield Digital Center for the History of Womens Education. Web. 10 May 2013. <http://greenfield.brynmawr.edu/items/show/2780>.
- [31] Mandell, Laura. "Special Issue: 'Scholarly Editing in the Twenty-First Century' A Conclusion." *Literature Compass* 7.2 (2010): 120-133. Web. 10 May 2013. <http://onlinelibrary.wiley.com/doi/10.1111/j.1741-4113.2009.00684.x/pdf>.
- [32] McGlothlin, James P., Latifur Khan, and Bhavani M. Thuraisingham. "RDFKB: A Semantic Web Knowledge Base." In proceeding of: IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16-22, 2011.
- [33] McPherson, Tara. "Why Are the Digital Humanities So White?, or, Thinking the Histories of Race and Computation." *Debates in the Digital Humanities*. Ed. Matthew K. Gold. Minneapolis: University of Minnesota Press, 2012. 139-60. Web. 10 May 2013. <http://journals2.scholarsportal.info/tmp/9027588419076974056.pdf>.
- [34] Mermin, Dorothy. *Godiva's Ride: Women of Letters in England, 1830-1880*. Indiana University Press, 1993. Print.
- [35] Michel, Jean-Baptiste, et al. "Quantitative analysis of culture using millions of digitized books." *Science* 331.6014 (2011): 176-182. Web. 10 May 2013. <http://www.sciencemag.org/content/331/6014/176>.
- [36] Moretti, Franco. "Graphs, Maps and Trees: Abstract Models for a Literary History." *New Left Review* 24 (2003): 67-93.
- [37] Musen, Mark A. "Tackling Climate Change: Unfinished Business from the Last 'Winter'." 11th International Semantic Web Conference. Boston, 2012.
- [38] Nicholson, Bob. "Counting Culture; or, How to Read Victorian Newspapers from a Distance." *Journal of Victorian Culture* 17.2 (2012): 238-246.
- [39] Oldman, Dominic . "The British Museum, CIDOC CRM and the Shaping of Knowledge." Blog post. 4 Sept. 2012. Web. 10 May 2013. <http://www.oldman.me.uk/blog/the-british-museum-cidoc-crm-and-the-shaping-of-knowledge>.
- [40] Olson, Hope A. *The Power to Name: Locating the Limits of Subject Representation in Libraries*. Dordrecht: Kluwer, 2002. Print.
- [41] Orlando Project. Web. 10 May 2013. <http://www.ualberta.ca/orlando>
- [42] Pinto, Sophia H., et al. Some Issues on Ontology Integration. Proceedings of the IJCAI-99 workshop on Ontologies and Problem-Solving Methods (KRR5) Stockholm, Sweden, August 2, 1999
- [43] Prins, Yopie. *Victorian Sappho*. Princeton University Press, 1999.
- [44] "Sapphism, n." OED Online. March 2013. Oxford University Press. Web. 10 May 2013. <http://www.oed.com/view/Entry/170877?redirectedFrom=sapphism>.
- [45] Stauffer, Andrew M. "Digital Scholarly Resources for the Study of Victorian Literature and Culture." *Victorian Literature and Culture*. 39.1 (2011): 293-303. Web. 10 May 2013.
- [46] Simpson, John, and Susan Brown. "LOD DOA? A Humanist Perspective on Building Ontologies." *Digital Humanities* 2013, Lincoln, Nebraska.
- [47] "Swoogle: Semantic Web Search Engine." Swoogle, 2007. Web. 10 May 2013. <http://swoogle.umbc.edu/>.
- [48] VIAF. OCLC, 2012. Web. 10 May 2013. <http://viaf.org/>.
- [49] Wiebel, Stuart. "Stuart Weibel Interviews Tim Berners-Lee." OCLC Research. <http://www.oclc.org/research/news/features/tbliview.html> Web.
- [50] Woolf, Virginia. *A Room of One's Own*. 1929. Harcourt, Brace & World, 1957. Print.
- [51] York, Lorraine Mary. *Rethinking Women's Collaborative Writing: Power, Difference, Property*. University of Toronto Press, 2002.
- [52] W3C. W3C Semantic Web Frequently Asked Questions. Accessed August 7, 2013. <http://www.w3.org/RDF/FAQ>