An Entity By Any Other Name: Linked Open Data as a Basis for a Decentred, Dynamic Scholarly Publishing Ecology

Susan Brown & John Simpson
University of Alberta

CWRC Project Team

INKE Research Group

Abstract

We propose linked open data as enabling a more interlinked and easily navigable scholarly environment that would permit: better integration of research materials with primary and secondary source objects and datasets; the potential to bridge but also address the specificities of the nomenclature, discourses, and methodologies of humanities disciplines and sub-disciplines; and the ability to respect institutional and individual investments in ownership or credit of resources by allowing for identifiable collections of data while fostering resource interlinking. Linked data can underwrite a publishing ecology based on collaborations between the scholarly, publishing, and library communities, but this vision is tempered by concerns about linked data publishing practices and infrastructure gaps with respect to enabling such collaboration, particularly in the humanities.

Susan Brown holds a Canada Research Chair in Collaborative Digital Scholarship at the University of Guelph, where she is a Professor in the School of English and Theatre Studies. Email: susan.brown @uoguelph.ca .

John Simpson is Compute Canada's Digital Humanities Specialist and has a diverse background in Philosophy and Computing. Email: john.simpson @computecanada.ca .

CWRC Project Team: the Canadian Writing Research Collaboratory is a digital humanities infrastructure project funded by the Canada Foundation for Innovation's Leading Edge Fund

Implementing New Knowledge Environments (INKE) is a Major Collaborative Research Initiatives research grant funded by the Social Sciences and Humanities Research Council of Canada.

Isolation from related materials plagues online scholarly digital resources within and beyond the humanities. A major complaint from scholars with respect to finding and using digital materials is that they exist in silos and are not interlinked with other relevant materials (Bulger, 2011; Frost & Dombrowski, 2011). This is equally true of conventional print-legacy publications such as online journals, scanned books, and e-books – access to which is often exacerbated by either paywalls or the database structures within which they are housed; digital humanities projects published on the Web by individuals or libraries; and mass digitization or aggregation projects. Indexing services help mitigate this problem somewhat, but meaningfully interconnecting resources with the materials they cite and the materials that cite them remains a challenge. Huge gains would result from being able to leverage and formalize, for instance, citation networks in our information environment, whether the resources in which the citations occur are formally or informally published, or whether they occur in the working annotations of individual scholarly users and the discourses that surround them in social media. At root, to use the discourse of entities or "things" associated with the Semantic Web or linked open data (LOD), this means being able to link various *entities* related to those resources to one another (WorldCat, 2015).

We have an unprecedented challenge and opportunity in the volume and variety of largely disconnected scholarly discourse circulating in digital form. Addressing this challenge in a feasible manner would do two important things. First, a higher level of interconnection and interoperability of text and contexts would go a long way in solving Gregory Crane's (2006) "million books" problem. It would enable scholarly inquiry to scale up in ways that have, to date, been accessible to only a very small proportion of humanities scholars with the funding and the skills to compile large datasets for their own use. Even those efforts have been inevitably limited by the fact that their datasets are, though large, nevertheless bounded. Second, interlinked and interconnected scholarly discourse stands a good chance of ratcheting up its impact, whereas that work is currently invisible to the major search engines and fails to register among the other sources of information that populate the Web. This is a particular cause for regret given both its relevance to many contemporary debates and its greater claims to authority and trustworthiness than many of those other sources.

This article takes up the smaller and more manageable problem of *interlinking* as a first crucial step toward interoperability in proposing linked open data, with its leveraging of entities and relationships, as a means of producing a more interconnected and more easily navigable knowledge environment. The fundamental building blocks exist to allow such a system to develop, and indeed key initiatives are underway within the library and museum communities and the publishing community. The focus here will be on the scholarly community and its ability to engage in these developments in ways that will both strengthen the overall shape of the Semantic Web and help the digital humanities overcome some major blockages that have been impeding its impact both within the traditional humanities and within the larger information environment. We do not employ an environmental metaphor – ecology – out of disregard for the tremendously detrimental global impacts of electronic waste and energy consumption (Digital Environmental Humanities, n.d.; Uddin & Rahman 2011; Widmer, Oswald-Krapf, Sinha-Khetriwal, Schnellmann, & Böni, 2005), nor to "cloud" the local features

and effects of what we are describing (Jaeger, Lin, Grimes, & Simmons, 2009). The metaphor of a publishing ecology highlights several aspects of the approach here.

As initially defined by Darwinian disciple Ernst Haeckel, ecology considers "the relations of the organism to the environment including, in the broad sense, all the 'conditions of existence'" (quoted in Egerton, 2013, p. 226). Applying an ecological framework stresses the extent to which any attempt to alter scholarly communications and discourse must be understood in terms of both diversity and systematicity, since it involves modifying the links between people and the material and institutional conditions in which they work. As Bonnie A. Nardi and Vicki O'Day (1999) argued in introducing the term:

> An information ecology is a complex *system* of parts and relationships. It exhibits *diversity* and experiences continual evolution. Different parts of an ecology *coevolve*, changing together according to the relationships in the system. Several *keystone species* necessary to the survival of the ecology are present. Information ecologies have a sense of *locality*. (n.p.)

Framing this as an ecological problem also allows us to think in terms of "ecotones," "an interface region between two different ecosystems" (Hegde, 2012) – dynamic regions where the mixing of populations at the margins of two different communities produces unusual pressures and stimulates change. This article identifies some of the characteristics of the ecotones associated with the edge zones between scholarly publishing and library communities; the citizen scholar, archival, and gallery and museum sectors would be worth similarly examining. Ecotones are understood to be crucial in supporting "diverse communities and … [affecting] the flow of materials across the landscape" (Risser, 1990, p. 9), which resonates with concerns surrounding the emergent Semantic Web (Brown & Simpson, 2013). Edge spaces are not vacant gaps, but fertile, if also conflictual, zones that are crucial to fostering a healthy and balanced information environment (Brown, 2011). Also relevant are the connotations of ecology as a social movement, the sense that there are better and worse ways of impacting an environment, and interventions should be beneficial in their long-term consequences beyond the immediate context.

Benefits of a linked open data knowledge ecology

So, then, how can linked data lead to a better publishing ecology for scholarship, and in particular allow scholarly publications to interact with, enhance, and ameliorate datasets being produced in the library and museum communities on the one hand and formal publishing ventures on the other? The focus here will be on several benefits that by no means exhaust the possibilities: 1) interlinking and, at least at the level of interface, integration of resources; 2) provision of context and relationship information as the foundation for a rich knowledge environment; 3) feedback loops that improve the quality of data, particularly that provided by large-scale information providers; and 4) incorporation of diversity of discourse, methodology, and data, including nuanced ontologies and datasets that respect the local and particular, including outliers that may appear as "noise" within large datasets.

1) The interlinking and, at least at the level of interface, integration of resources

This is the preeminent or umbrella use case for linked open data (LOD) applications in fields related to the humanities. As Jim Hendler (2011) contends, the Semantic Web's Resource Description Framework (RDF) got right what Extensible Markup Language (XML) got wrong: external linking. Much energy is currently focused on the potential for LOD to help in the exposure and integration of large datasets. The library and museum communities are the areas where these sorts of initiatives are most prominent, with initiatives such as the Europeana LOD data release and pilot projects (see Europeana Labs, 2015), and the British Museum Collection of RDF (datahub, n.d.). Closer to home is the "Out of the Trenches" proof of concept developed by the Pan-Canadian Documentary Heritage Network (PCDHN, n.d.), including major research libraries and Canadiana (Wuppleman, 2012), and more recently the innovative Muninn Project that uses linked open data to produce simulations of WWI trenches (Muninn Project, n.d.; Warren, 2012). The current Linked Data for Libraries initiative and VIVO project in the United States are also using linked data to aggregate scholarly data and library holdings, leveraging open library resources such as the Virtual International Authority File (LD4L, 2014; VIVO Open Research Networking Community Group, 2015).

All of these projects present strong use cases for the use of linked data to expose and interlink research results and researcher publication networks. None of them build scholarly research activity into their vision of the resulting publishing ecology. The Online Computer Library Center (OCLC) has done some work on collaborating with scholars in its linked data initiatives (Klein, 2012a), but has also acknowledged significant stumbling blocks in such collaborations. Apparently it has a more established and indeed automated collaboration with the Wikipedia community (Klein, 2012b; OCLC Research, 2014; Smith-Yoshimura, Michelson, & Mardutho, 2013). Although there are certainly some exceptions – for instance the DM2E or Digitised Manuscripts to Europeana project, which is connected with the Digital Research Infrastructure for the Arts and Humanities (DARIAH-EU) infrastructure initiative – active scholarly research projects are being omitted from the process and workflows involved in producing and publishing large datasets of humanities objects. Omitting the participation of active scholars and the interlinking of active research projects, even though it would necessitate a departure from print-oriented understandings of resource stability and the boundaries of archives, seems like a missed opportunity to enrich these resources further.

2) Provision of context and relationship information as the basis for a rich knowledge environment

Given the high expectations of currency from Web resources, interlinking scholarly research materials with publishing datasets would provide valuable contextual information for those datasets, since scholarly work relates primary sources and published scholarship with debates of contemporary relevance. As one information scientist, R.J. Searle, put it, humanists, in a sense, "are curators par excellence of scholarly information" because they transform primary "raw" data into secondary "institutional" content (quoted in Benardou, Constantopoulos, Dallas, & Gavrillis, 2010, p. 28). Much stands to be gained from the better integration of research materials with the primary

and secondary source materials on which they draw. Beyond linking to external resources for contextual information, emerging standards like the Open Annotation Data Model (2013) offer the potential for online editions of primary literary texts, for instance, to draw on research notes produced by scholars in other contexts.

3) Feedback loops that improve the quality of data, particularly that provided by large-scale information providers

Scholars have the expertise and motivation to correct the dirty data that is out there. Some groundbreaking projects are building bridges between large-scale digital content providers and the scholarly community to mutual benefit (e.g., eMOP, n.d.). Such endeavours can channel the scholarly itch to correct errors into the enhancement of large-scale digitization efforts, by enabling users to correct optical character recognition (OCR) errors, or note faultily scanned images embedded in collections. What is needed are tools to allow the data providers to easily harvest back information about corrections into their source datasets, to aggregate this information into interfaces with provisions for filtering such information by provenance and trust criteria, and to incorporate the results via machine learning back into OCR processes to improve overall accuracy.

4) Incorporation of diversity of discourse and methodology and data

The humanities have a great deal to add to the development of a larger linked data ecology in the area of nuanced ontologies and datasets that respect the local and particular, including outliers that may appear as "noise" within large datasets. The potential to address the specificities of the nomenclature, discourses, and methodologies of humanities disciplines and sub-disciplines while also bridging them, and the ability to respect institutional and individual investments in ownership or credit of resources by allowing for identifiable collections of data while also fostering resource interlinking, will counter tendencies of linked open data to occlude difference and diversity as a result of the process of scaling up.

Modelling an open ecology

As a starting point, we here propose a very high-level model for a decentred, dynamic publishing ecology based on collaborations among scholarly, publishing, and library communities founded in linked data principles (see Figure 1).

The solid coloured lines between the rough categories of content are meant to represent the high degree of complementarity in the data held, and the ability of each domain to enhance the other in a range of ways. It is suggestive rather than comprehensive. Each of the domains is only minimally contained in a porous cloudlike shape that overlaps with the others, and above them are the linked data services that are essential to a dynamic and productive ecology of the kind envisioned here. The broken arrows moving into the ecotones between the domains illustrate the extent to which the synergies indicated by the solid arrows presuppose such services, but they are not yet available.

Figure 1: Sketch of LOD-based dynamic scholarly publishing ecology
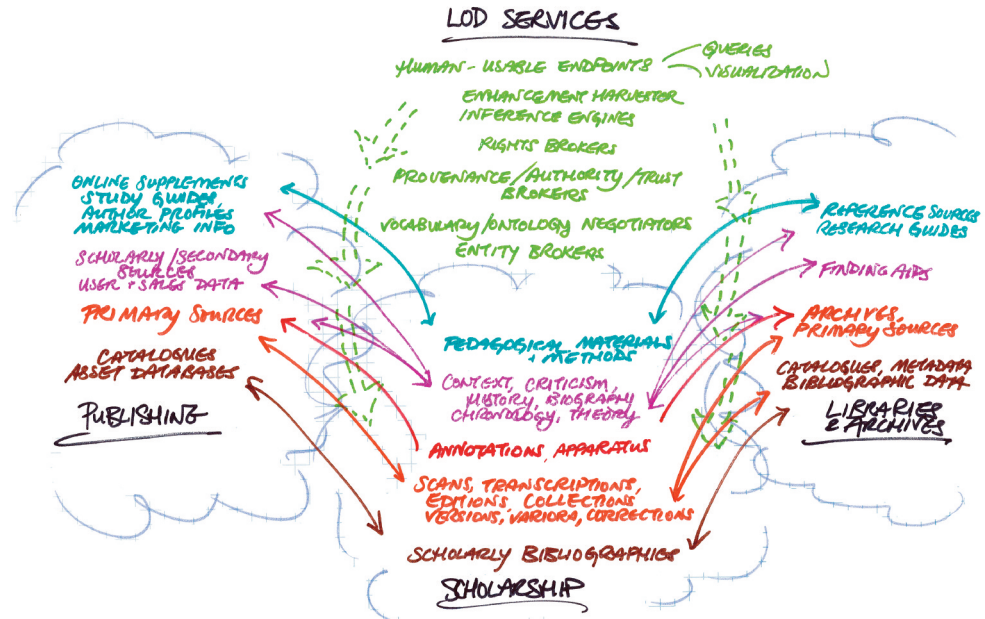
FUNCTIONALITY GAPS

As the broken arrows indicate, the vision of the glory the Semantic Web might offer must be tempered by a consideration of the current state of linked data publishing practices and infrastructure. There are significant gaps in tools and infrastructure that need to be filled before this model could become a reality. We focus here on two complementary gaps in the LOD publishing ecology with respect to refining entities and nuancing the ontologies that interrelate them.

ENTITY DISAMBIGUATION/ALIGNMENT/LINKAGE

Fully automated conversion or aggregation of existing materials into LOD produces results that erase distinctions and differences around which much work in the humanities revolves. Refusal of automated processing may be why humanities "linked" datasets are frequently self-referential, with few or no links to external data. An urgent need exists for LOD technologies that allow efficient human oversight, refinement, and correction of automated processes in order to ensure that humanists can create or adapt linked datasets in which they have confidence. What is required is a workflow that allows researchers to take an existing structured or unstructured dataset and perform a series of operations to prepare it as LOD. The operations are as follows: 1) perform named entity and triple recognition/extraction on the dataset, which may involve using training sets to obtain accurate results; 2) match the results to existing LOD collections that will be user selectable/configurable; 3) present users with candidate matches for ambiguous entities and triples so as to allow them to process imperfect matches and triple candidates; 4) based on this input, produce LOD annotations of the data and/or embed LOD identifiers in the data (crucial for humanities projects with embedded metadata), drawing on the Open Annotation Data Model (2013); and 5) feed the results back in a machine learning system to improve future matching.

Susan Brown, John Simpson, CWRC Project Team, & INKE Research Group. (2015). An Entity By Any Other Name: Linked Open Data as a Basis for a Decentred, Dynamic Scholarly Publishing Ecology. *Scholarly and Research Communication, 6*(2): 0201212, 11 pp.

Open source components for such a workflow exist in tools such as the Stanford Named Entity Tagger (n.d.) and LODE (n.d.), the Linked Open Data Enhancer developed in partnership with the Indiana Philosophy Ontology (InPho) Project (2013). What does not exist is a usable and accessible workflow that could serve a wide range of types of texts. Such a workflow would advance a number of existing scholarly LOD projects. It would fill a major infrastructure gap to enable the interlinking of publishing, library, and museum data with scholarly data to create a richly symbiotic set of relationships. Beyond this, such a workflow would encourage the use of LOD by humanists, pushing humanities data to new levels of interoperability while enhancing existing datasets, and allowing for new kinds of inquiry and inferencing across cultural datasets. The lack of such a tool is also felt by major information providers. Organizations such as the Library of Congress and OCLC, the nonprofit Online Computer Library Center that hosts WorldCat, which provide the ultimate authority datasets in our field, will be looked to for disambiguating linked data entities, but their production of linked data is hampered by the lack of the kind of processes described here. For instance, OCLC will soon release approximately 100 million personal names as linked data, in addition to the existing names and 197 million titles of works already released. However, to generate this dataset, OCLC has opted to ignore imperfect matches; for instance, authors with slight variations in the representation of their names (e.g., "E. Pauline Johnson" versus "Pauline Johnson"), will not be understood as the same entity (Fons, 2014). At the Coalition for Networked Information meeting in the fall of 2014, the principals of large research-oriented linked open data projects agreed that reconciliation services are urgently required, and yet no one in that community has undertaken to produce such a tool.

Although relatively modest and quite feasible, a usable and generalized workflow of this type could be a game changer. As Dominic Lam (2014) argues, such workflows are crucial to scaling up digital humanities research. Moreover, as Semantic Web technologies become more pervasive (as with Google), the public impact of exposing and interlinking large bodies of humanities data may be considerable.

Navigating between ontologies

A survey we have done of the implementation of ontologies on the Semantic Web shows that the graph of ontology usage has a very long tail, suggesting that more convergence in ontology adoption is needed if the aim is an interconnected Web (Simpson, Brown, & Goddard, 2013). The flexibility of linked data technology lies in the fact that each datastore can develop its own vocabulary and ontology to suit its needs, and yet link out to other datastores. However, linking up with other data means connecting one ontology to another, and this brings with it a pressure toward generalization rather than specificity. It is no accident that the most commonly used RDF vocabulary is the Dublin Core Metadata Initiative (n.d.), the success of which can be attributed in large part to its great simplicity and very broad applicability (Simpson et al., 2013). Yet generalization makes data much less useful for humanities inquiry, enabling "information jukeboxes" (Oldman, 2012) rather than nuanced research tools. Initiatives such as Linked Data for Libraries (LD4L) (n.d.) are going with major ontologies such as schema.org and Friend of a Friend (FOAF) in order to ensure exposure through the large search engines. While this is in itself a logical and laudable

goal, it means compromise, including misrepresenting and/or de-specifying some of the features of ontologies developed specifically for bibliographical data in order to make it "fit" the dominant ontology (Krafft & Cramer, 2014). If such standards occlude even the fairly straightforward categories of major cataloguing standards, how much will be lost of the eclectic, the nuanced, and the more precise features in humanities linked open datasets when it comes to aligning ontologies?

What is required is a tool set for linked data access to help researchers and information specialists select datasets, identify significant differences within and between them, and navigate those differences according to the particular methodological needs of their inquiry. The tool would permit the bridging of entire data sets of the user's choice and enable control of how RDF ontologies are mobilized and subsequently how the inferences made. Bridging scholarly repositories such that they retain some of the richness of their local ontologies is key to guarding against over-generalization in Semantic Web ontologies. Consider a researcher interested in exploring the complicated and unsettled question of women writers' use of pseudonyms and their relation to reception history. She might work with data from a number of existing research datasets on women's writing, all of which contain rich reception content and highly detailed information on pseudonyms. The tool would allow her to see how these collections' ontologies compared to those of more general datasets like the Virtual International Authority File and DBpedia, noting differences in the treatment of personal names. She could decline to move to a common denominator by flattening all types of names into a "creator" role, electing to retain greater granularity in the data models associated with the research collections. The tool would afford ways of "narrowing up" by leveraging more precise relations to inform more general ones. Her decisions would be informed by the ability to select sample entities for authors she knows and view the consequences of her choices in the output data, which would group materials or infer triples differently depending on the researcher's choice. A researcher's choices could be saved into the tool's library, for later use by her or others, to document her research process. This kind of specialist engagement with ontologies could conceivably work against the homogenizing tendencies of the Semantic Web, if a feedback loop could be created to harvest the results of trusted work so as to respecify relationships that have been overgeneralized in the production of the linked data, or enrich with greater specificity datasets that were not precise or complete at the outset.

Conclusion

This discussion by no means exhausts the gaps. The model indicates a range of LOD services that are needed, most of which do not yet exist at all or, at least, in the mature and generalized form needed to support the kind of dynamic interchange of LOD envisioned here. They include the need for better mechanisms for establishing automated conditions for evaluating the provenance, authority, and trustworthiness of LOD resources, and for tools to harvest and incorporate corrections and enhancements. Rights are of course a major consideration. There remains also the fact that despite some nice bespoke interfaces tailored to specific collections, we lack really good human-usable interfaces for the Semantic Web at large, whether these are for queries that draw on the semantic structure or visualizations of portions of the graph. We highlight here two gaps that we consider particularly significant for the humanities

community. The lowest hanging fruit for work in this area lies in entity identification and linking, which will allow humanities data to move onto the Semantic Web and constitute a major component of a public-facing humanities. An ontology negotiation tool, or what we like to think of as a "difference engine" (in homage to Charles Babbage), might be the most significant contribution that the humanities could make to the emerging Semantic Web publishing ecology, particularly if it is able to enrich ontologies in other areas such as the publishing and library sectors. An entity-based approach to digital scholarly publishing allows for the incorporation of living scholarship alongside print-like resources, reflecting the increasingly dynamic nature of scholarly production in the digital age as a necessary component of the online knowledge environment. It offers digital scholars local solutions with respect to authority control, information retrieval, information visualization, and in the longer term inference and reasoning that draws on other knowledge sources. In short, it represents an opportunity for fruitful collaboration with other closely related sectors of the knowledge economy, combined with the potential to influence the Web more directly as an evolving space of knowledge production and dissemination.

Websites

DBpedia, http://wiki.dbpedia.org

Schema.org, http://schema.org

Virtual International Authority File, https://viaf.org/

References

Benardou, Agiatis, Constantopoulos, Panos, Dallas, Costis, & Gavrillis, Dimitris. (2010). Understanding the information requirements of arts and humanities scholarship. *International Journal of Digital Curation*, 5(1), 18-33.

Brown, Susan. (2011). Don't mind the gap: Evolving digital modes of scholarly production across the digital-humanities divide. In Daniel Coleman & Smaro Kamboureli (Eds.), *Retooling the humanities: The culture of research in Canadian universities* (pp. 203-231). Edmonton, AB: University of Alberta Press. 203-231. URL: http://hdl.handle.net/10402/era.25382 [July 17, 2015].

Brown, Susan, & Simpson, John. (2013). The curious identity of Michael Field and its implications for humanities research with the semantic web. *IEEE International Conference on Big Data 2013*, 77-85. URL: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6691674&tag=1 [October 9, 2015].

Bulger, Monica E., Meyer, Eric, De la Flor, Grace, Terras, Melisa, Wyatt, Sally, Jirotka, Marina, Eccles, Katherine, & Madsen, Christine McCarthy. (2011). Reinventing research? Information practices in the humanities. *Information Practices in the Humanities, March. A Research Information Network Report*.

Crane, Gregory. (2006). What do you do with a million books? *D-Lib Magazine, 12*(3).

Datahub. (n.d.). *British museum collection*. URL: http://datahub.io/dataset/british-museum-collection [July 17, 2015].

Digital Environmental Humanities. (n.d.). *Home*. URL: http://dig-eh.org/ [11 September 2015].

Dublin Core Metadata Initiative. (n.d.). *Dublin core metadata initiative.* URL: http://dublincore.org/ [July 17, 2015].

Egerton, F.N. (2013). History of ecological sciences, part 47: Ernst Haeckel's ecology. *Bulletin of the Ecological Society of America*, 94(3), 222-244.

eMOP: Early Modern OCR Project. (n.d.). *Home*. URL: http://emop.tamu.edu/ [July 17, 2015].

Susan Brown, John Simpson, CWRC Project Team, & INKE Research Group. (2015). An Entity By Any Other Name: Linked Open Data as a Basis for a Decentred, Dynamic Scholarly Publishing Ecology. *Scholarly and Research Communication, 6*(2): 0201212, 11 pp.

9

Europeana Labs. (2015). *Europeana linked open data*. URL: http://labs.europeana.eu/api/linked-open-data/introduction/ [July 17, 2015].

Fons, Ted. (2014). *Transforming bibliographic records into linked open data (LOD)*. Panel presentation at the Coalition for Networked Information Fall 2014. URL: http://www.cni.org/topics /information-access-retrieval/exposing-library-collections-on-the-web-challenges-and-lessons-learned/ [July 17, 2015].

Frost Davis, Rebecca, & Dombrowski, Quinn. (2011). Divided and conquered: How multivarious isolation is suppressing digital humanities scholarship. Washington, DC: National Institute for Technology in Liberal Education. URL: https://web.archive.org/web/20130927214037 /http://www.nitle.org/live/files/36-divided-and-conquered [July 17, 2015].

Jaeger, Paul T., Lin, Jimmy, Grimes, Justin M., & Simmons, Shannon N. (2009). Where is the cloud? Geography, economics, environment, and jurisdiction in cloud computing. *First Monday*, *14*(5).

Hegde, Medha. (2012). Ecotones: the transitional zones. *Biotech Articles*, *12*. URL: http://www .biotecharticles.com/Biology-Article/Ecotones-The-Transitional-Zones-2191.html [July 17, 2015].

Hendler, Jim. (2011). *Why the Semantic Web will never work*. Presented at the 8th Extended Semantic Web Conference (ESWC) in Heraklion, Greece. URL: http://videolectures.net/eswc2011_hendler _work/ [July 17, 2015].

Indiana Philosophy Ontology (InPho) Project. (2013). *The InPho Project.* URL: https://inpho.cogs .indiana.edu/ [July 17, 2015].

Klein, Max. (2012a). OCLC makes new moves connecting Wikipedia and libraries. *OCLC eNews*, *31*. URL: http://www.oclc.org/en-europe/publications/newsletters/enews/2012/31/en-08.html [July 17, 2015].

Klein, Max. (2012b). *VIAFbot debriefing*. OCLC Research. URL: http://hangingtogether.org/?p=2306 [July 17, 2015].

Krafft, Dean, & Cramer, Tom. (2014). *Video: Linked data for libraries (LD4L) project update*. Coalition for Networked Information. URL: http://www.cni.org/news/video-linked-data-for-libraries-ld4l-project-update/ [July 17, 2015].

Lam, Dominic. (2014, September 8). Big data challenges in social sciences & humanities research. *Datanami*. URL: http://www.datanami.com/2014/09/08/big-data-challenges-social-sciences-humanities-research/ [July 17, 2015].

LD4L: Linked Data for Libraries. (2014). *Duraspace*. URL: https://wiki.duraspace.org/pages /viewpage.action?pageId=41354028 [July 17, 2015].

LODE: Linked Open Data Enhancer. (n.d.). URL: http://www.linkedhumanities.com [July 17, 2015].

Muninn Project. (n.d.). URL: http://blog.muninn-project.org/ [July 17, 2015].

Nardi, Bonnie A., & O'Day, Vicki L. (1999). Chapter 4: Information ecologies. *First Monday*, *4*(5). URL: http://firstmonday.org/ojs/index.php/fm/article/view/672/582 [July 17, 2015].

OCLC Research. (2014, March 3). *Scholars' contributions to VIAF*. URL: http://oclc.org/research /activities/viaf-scholars.html [July 17, 2015].

Oldman, Dominic. (2012, September 4). The British museum, CIDOC CRM and the shaping of knowledge. *My Blog: The Blog of Dominic Oldman*. URL: http://www.oldman.me.uk/blog/the-british-museum-cidoc-crm-and-the-shaping-of-knowledge/ [July 17, 2015].

Open Annotation Data Model. (2013). *W3C*. URL: http://www.openannotation.org/spec/core/ [July 17, 2015].

Pan-Canadian Documentary Heritage Network. (n.d.) *Linked Open Data (LOD) Visualization 'Proof-of-Concept.'* URL: http://www.canadiana.ca/sites/pub.canadiana.ca/files/PCDHN%20Proof-of-concept_Final-Report-ENG_0.pdf [13 September 2015].

Risser, Paul G. (1990). The ecological importance of land-water ecotones. In H. Décamps & R.J. Naiman (Eds.), *The ecology and management of aquatic-terrestrial ecotones* (pp. 7-21). Paris, FR: UNESCO.

Searle, John R. (1995). *The construction of social reality.* New York, NY: Simon and Schuster.

Simpson, John, Brown, Susan, & Goddard, Lisa. (2013). A humanist perspective on building ontologies in theory and practice. *Digital Humanities Conference Abstracts 2013.* Lincoln, NE: University of Nebraska. URL: http://dh2013.unl.edu/abstracts/ab-413.html [July 17, 2015].

Smith-Yoshimura, Karen, Michelson, David, & Mardutho, Beth. (2013, March 27). Irreconcilable differences? Name authority control & humanities scholarship. *OCLC Research.* URL: http://hangingtogether.org/?p=2621 [July 17, 2015].

Stanford Named Entity Tagger. (n.d.). URL: http://nlp.stanford.edu:8080/ner/ [July 17, 2015].

Uddin, Mueen, & Rahman, Azizah Abdul. (2011). Techniques to implement in green data centres to achieve energy efficiency and reduce global warming effects. *International Journal of Global Warming, 3*(4), 372-389.

VIVO Open Research Group. (n.d.) *W3C.* URL: http://www.w3.org/community/vivo/ [July 17, 2015].

Warren, Robert. (2012). Creating specialized ontologies using Wikipedia: The Muninn experience. Berlin, DE: Proceedings of Wikipedia Academy: Research and Free Knowledge (WPAC2012). URL: http://hangingtogether.org/?p=2306http://hangingtogether.org/?p=2306 [July 17, 2015].

Widmer, Rolf, Oswald-Krapf, Heidi, Sinha-Khetriwal, Deepali, Schnellmann, Max, & Böni, Heinz. (2005). Global perspectives on e-waste. *Environmental Impact Assessment Review, 25*(5), 436-458.

WorldCat Entities. (2015). *OCLC Developer Network.* URL: http://oclc.org/developer/develop/linked-data/worldcat-entities.en.html [July 17, 2015].

Wuppleman, William. (2012, November 7). Out of the trenches: A linked open data project. *Canadiana.* URL: http://www.canadiana.ca/en/pcdhn-lod [July 17, 2015].