The Performativity of Data: Re-conceptualizing the Web of Data

Marie Joan Kristine Gloria
Rensselaer Polytechnic Institute
Troy, NY
glorim@rpi.edu

Dominic DiFranzo
Rensselaer Polytechnic Institute
Troy, NY
difrad@rpi.edu

Marco Fernando Navarro
Rensselaer Polytechnic Institute
Troy, NY
navarm@rpi.edu

Jim Hendler
Rensselaer Polytechnic Institute
Troy, NY
hendler@cs.rpi.edu

ABSTRACT

As the discipline of Web Science matures, its interdisciplinary claim has many researchers unsure about its core theory and methodology. Instead, we often see results that are more multi-disciplinary than interdisciplinary. The following contribution attempts to recast our understanding of the current methodologies and tools leveraged by the Web Science community. Specifically, we review the Semantic Web and Linked Data technologies not just from a technical perspective; but, through a critical reading of key social theories such as Goffman's theory of performance. Our goal is to re-conceptualize the performativity of semantic web tools, their boundaries, and any potential avenues for future research.

Author Keywords

Experimental methods, Semantic Web, Web Science Theory

ACM Classification Keywords

H.1.1. Systems and Information Theory (E.4)

General Terms

Theory; Human Factors

INTRODUCTION

As the Internet continues to provide both an object of study and research tools, it raises many questions for methodologies of Web Science research. We now live in an era where big data, abundant data, and accessible data exists and where relational information is its most relevant characteristic. The challenge is no longer quantitative [22]. Instead, we must find new ways to identify, refine and contextualize data [22]. For Web Science, mastering data to scale while grounding it in viable social theory remains disjointed.

The Semantic Web and Linked Data are tools that have the potential to realize this. Specifically, these tools enable automated querying, sharing and interpretation of the data from distributed heterogeneous formats [28]. The World Wide Web Consortium (W3C) also defines the Semantic Web as "a common framework that allows data to be shared and reused across application, enterprise, and community boundaries" [37]. The advantage is seamless integration and the addition of contextualized information of a dataset, which most importantly is machine readable. This vision of the Web of Data moves us beyond unstructured "throw-away" data to a more sophisticated system that attempts to bridge the worlds of knowledge representation and information presentation [36].

However, we have yet to fully uncover the potential of merging different methodological approaches to and interpretations of this data; particularly for Web Science research. As we mentioned, experts and critics share concerns of its epistemological, political and ethical effects of data. Thus, we move from big data to broad data where we must re-examine the types of questions we ask of the data and exactly how to ask these questions [16]. Specifically, we consider the methodological advantages and disadvantages of leveraging mixed methods and new research tools. We also begin to discuss technologies in the context of traditional human social theory by examining the performativity of these tools and re-conceptualizing boundaries and possible avenues for future research. For as Pickering noted, "it [the performative idiom] precipitates instead all sorts of fascinating and important lines of inquiry, from the narrowly academic to the globally political" [30].

WebSci'13, May 2–4, 2013, Paris, France.

This paper just scratches the surface and is by no means exhaustive. It is not within the scope of this paper to dismiss or replace long-standing theories of human social behavior. Instead, we propose the opposite. We argue that the Semantic Web gives us the opportunity to apply theories of human social behavior and interaction to web practices. For, unlike the World Wide Web, the Semantic Web offers a more complete graph of knowledge like the history of and revisions to certain datasets. This convergence moves the human into encoded human practices creating the need for a future "semantic web theory" that is both innately human but also something more. We hypothesize that as we continue down this data-centric, Web-involved path, it is imperative that we consider how research techniques must sensibly evolve over time.

RESEARCHING (BIG) DATA

In the late 1890s, renowned mathematician, Pierre-Simon Laplace wrote, "this simplicity of ratios will not appear astonishing if we consider that all the effects of nature are only mathematical results of a small number of immutable laws" [17]. Laplace extended his work in probability theory into human behavior, which he conjectured could be distilled down to statistical, predictable outcomes [19]. Early entrances into social research driven by big data have arguably inched closer to Laplace's thesis. For the humanities and the social sciences, the promise of big data is one that makes social spaces more quantifiable [5].

However, a growing number of researchers now caution against this data-centric analysis for reliably predicting human behavior or answering complicated large-scale human behavioral problems. Latour noted:

> "All too often, fields have been divided between number crunching, devoid (its enemies claim) of any subtlety; and rich, thick, local descriptions, devoid (its enemies say) of any way to generalize from these observations. Many domains have abandoned the hope of proving any point by transforming quantities into qualities, and qualities into quantities". [19]

Moreover, the sheer volume of available data complicates varying methodological philosophies on a more data logic level. Researchers must then consider broadening their scope of observation and to contract more inductive reasoning. No matter the methodological approach, the collection of data must be rigorous, disciplined, analyzed and systematically presented [9]. The incorporation of the Web both as the subject of study and as a research tool complicates this process even further. *This blurring challenges traditional boundary settings in research; and thus, requires a rethinking of both methodological frameworks and theory.* More importantly, as Web Science touts its multi-discipline nature, we must too ask what can be discovered and what can be understood at the intersection of computer analysis of social and cultural data versus just traditional qualitative methods or vice versa [22].

LINKED DATA AND THE SEMANTIC WEB

The Semantic Web was first introduced in a 2001 *Scientific American* issue. In this article, a majority of the focus was built around the idea of software agents that would make decisions and interact on the Web in much of the same way that people do. As Semantic Web technologies grew and matured over time, focus shifted from software agents to linked data and vocabulary usage. The linking of different heterogeneous datasets and structures together in a more bottom-up, decentralized and web-like way has become one of the biggest advantages of the Semantic Web. Hall noted:

> "It is hoped that the Semantic Web will exhibit the same network effects that promoted the growth of the WWW. Just as using the Web became more valuable to more people as Web usage increased, the more that people share data, which can be mapped onto URIs and linked to other data, the more valuable that data is" [15].

Unlike the World Wide Web, the Semantic Web explicitly expresses the concepts, ideas, information and all the different relationships they can have among each other, which is denoted by Universal Resource Identifiers (URIs; a superset of URLs). URIs guarantee uniqueness in naming concepts, ideas and relationships in the Semantic Web. Such features make it much more manageable to use this information in more automatic and quantifiable ways. More importantly, when data is encoded to be more explicit, hidden data emerges and can be used in new and interesting ways.

In addition, the Semantic Web's Resource Description Framework, or RDF, allows for a simple way to model information and data as graphs. RDFs are denoted using URIs naming both the nodes and the edges of the graph. The triple consists of three parts: the *Subject* (the main concept we are describing, or the starting node that we are referring to), the *Predicate* (the edge or named relationship coming from the Subject to another node), and the *Object* (the node that the Subject is linking to with the Predicate). Semantic Web practitioners use URIs to name the three parts of the triple. For example: "California is part of the United States." Here, "California" is the Subject; "is part of" is the Predicate, and "United States" is the Object. The RDF notation of the "California" example may then look like this: *http://example.com/California http://example.com/isPartOf http://example.com/USA*. By using many triples, researchers can begin to construct and describe a very complex graph. By modeling data as a graph, RDF allows for easier linking and mashing together of heterogeneous datasets by explicitly connecting common nodes together.

This process creates a broader, more complex and complete graph of data and knowledge representation that can be reused and queried as a whole. This idea of linking common ideas and concepts together to form a more complete graph is referred to as "Linked Open Data". Linked Open Data consists of many different types of datasets joined together in a complex graph that may reuse preexisting vocabularies to help describe the relationships in the particular dataset. For example, the Tetherless World Constellation (TWC) at Rensselaer Polytechnic Institute (RPI) developed a Semantic Web-based portal to support the deployment of Linked Open Government Data (LOGD). The LOGD portal is an open source infrastructure supporting linked open government data production, consumption, management, and visualization. Different government agencies and organizations supply the datasets, which all retain their own degree of authority, policies, documentation, and trustworthiness. Additionally, the portal captures and explicitly reveals the workflow and history of the data. The LOGD Portal and its semantic framework begin to address some of the concerns voiced by critics of empirical driven social studies. For example, the system encodes not just the data; but the practice by which the data is collected and changed, mitigating concerns of authenticity.

We argue that given the graph-based and networked nature of the Web, the tools used to research and collect data from it should mirror its structure. Then, we can better recognize patterns and connections that may not be explicit at the beginning.

CONVERGING THEORY AND TECHNOLOGY

"Change the instruments, and you will change the entire social theory that goes with them" [17]. Understanding the technicalities is certainly one-half of the puzzle. The other half re-conceptualizes this convergence of technology and theory. Influenced by the work of Landow, we move away from the technical explanation to discuss the parallels between semantic web technologies and critical theory [17]. Additional authors such as Feenberg, Bolter, Lanham, etc., have also explored the nexus of technology and critical theory questioning linearity, marginality, centrality and the convergence to a more graph-like, networked world. Landow's scholarly work illuminated a shift in contemporary episteme where cultural, literary and computational theories were beginning to converge. Landow wrote:

> "The parallels between computer hypertext and critical theory have many points of interest, the most important of which, perhaps, lies in the fact that critical theory promises to theorize hypertext and hypertext promises to embody and thereby test aspects of theory". [17]

As we now shift towards a more structured knowledge representation environment, we build upon this literature. The Semantic Web and Linked Open Data, like Landow's hypertext example, warrants the same type of examination. Only this time, the tools call for a review on a data logic level. Applicable to this paper are Goffman's region theory, as a way of thinking about the separations (both physical and metaphorical) between researcher, data, and data sets; and, Burke's concept of the terministic screens as a means of viewing and interpreting human communicative practices vis-à-vis the Semantic Web. It is to our knowledge that no other attempt of such a reading has been made in context to the Semantic Web or Linked Open Data.

In beginning to build a conceptualization of a Semantic Web performance, it is useful to utilize Goffman's definition of a performance as an "arrangement which transforms an individual into a stage performer, the latter, in turn, being an object that can be looked at in the round and at length without offense, and looked to for engaging behavior, by persons in an 'audience' role" [13]. This definition, which makes affordances for a performer being transformed into an object, can be placed over the Semantic Web as we know it to be without much of a stretch. The objectification of human acts on the Web (under which data can be linked, organized, labeled, associated and/or "read") is thus "performed" on the Semantic Web. Arroyo reinforces this point with her description of the World Wide Web as "not a thing nor a verb or a noun" [2]. Instead, the Web's evolution, which we consider to include the Semantic Web, manifests a performance between actors from society and the Web's architecture.

However, it is our opinion that rather than defining players in the Semantic Web as actors or audience members, situating the notion of human-computer activity as a theatrical performance within Goffman's theory of region and region behavior is more effective [12]. Wherein the regions in the Semantic Web become specific, delineated, yet immensely interactional human communication paradigms. The following will present the following key points:

- Goffman's discussion of boundary setting and how this is applicable to the Semantic Web.
- How these boundaries establish a "backstage" and "front-stage", which gives way to flow of information and performances.
- How multiple backstages are constructed by utilizing semantic technologies.
- How semantic-enhancements and the use of provenance makes data "visible", affecting the flow of information between the back region and front region, which in turn, complicates our traditional understanding of the researcher and subject relationship.
- How Burke's concept of terministic screens allows us to view, read, and make sense of data as it is performed on the Web and viewed through the use of Semantic Web tools.

Goffman, in writing about the boundaries that exist among participants and parties, specifically criticizes regions and how they inform and explain boundaries. He wrote:

> "A region may be defined as any place that is bounded to some degree by barriers to perception. Regions vary, of course, in the degree to which they are bounded and according to the media of communication in which the barriers to perception occur". [12]

How do these physical boundaries, in Goffman's examples, correlate to the assumed boundaries (if even they should be denoted as such) of the Semantic Web? We propose that the Semantic Web, in its ability to make visible a seemingly endless quantity of data, should nonetheless be considered to occur within specific regions; and, it is through regions that the Semantic Web is denoted: through the "subject, predicate, object-structure" that the Semantic Web is built upon. As was noted earlier, the Semantic Web is unlike the World Wide Web in that it graphs concepts, ideas, and information as well as the different relationships among identifiers. Through the use of URIs, the Semantic Web names each of these human acts and performances with an identifiable symbol that makes data and relationships among the data uniquely quantifiable.

If we conceptualize the Semantic Web as a performance containing multi-foci, we can begin to more accurately conceptualize the Semantic Web graph. Again, we emphasize that this is only one conceptualization of the Semantic Web and that this conceptualization will be used in coordination with other conceptualizations, which will possibly show a more complete and nuanced Semantic Web. If we apply Goffman's concept of regions to the Semantic Web, we can organize the multitudinous nature of broad data into distinct regions and theorize about the relationships about and among regions.

To illustrate this further, we consider Goffman's notion of "impression management" specifically within the context of observing the performer leaving "the back region and enter[ing] the place where the audience is to be found...for at these moments one can detect a wonderful putting on and taking off of character" [12]. To emphasize, Goffman noted that vital secrets of a show are visible in the backstage and that performers behave out of character while there. Thus, performers while in the backstage, undergo a process of mediating "deterministic demands that surround them" [12]. This conceptualization of activity that is "performed" in the backstage, gives us an effective way to view web users, especially when an actor believes that his/her performances are not being viewed, monitored, tracked or recorded. We see such examples in media coverage of companies, like Google, who can accurately predict user search behaviors online.

It is natural to expect that the passage from the front region to the back(stage) region will be kept closed to members of the audience or that the entire back region will be kept hidden from them [12]. For traditional research models, this comparison holds true in that subjects (in this case, Web users and data-generators) are often kept from fully understanding—or seeing—the details of the "experiments" they unknowingly (or knowingly) participate in. And yet, it is precisely the existence of the Semantic Web, and semantic web tools, that makes all of this performative data "visible."

The Semantic Web reconciles this interplay of data. First, the Semantic Web exhibits multiple backstages, which are best represented through the application of discourses. For example, we may reflect on: *what is in front of the researcher becomes the backstage* - the underpinning of her research subject; which, to the researchee, is similarly another backstage, supposedly only visible to him. In this "revealing", the backstages both researchers and users are privy to (and engaged with in varying contexts) are linked to various backstage processes such as workflows and user-generated patterns (which, in a traditional stage performance, would be difficult to reveal to one other). The result of these linkages is intercontextual; highlighting contexts of the data like how it was created, by whom, and for what purpose. In this way, the Semantic Web unlocks data for researchers to understand and use this data better and more efficiently.

On yet another level, Goffman's performance theory also enables us to complicate the impressions made by both the researcher and subject as they perform towards each other on the Semantic Web. This is particularly poignant as criticism of big data research questions the act of interpretation of data [5]. There is, as Butler writes, a "normative force of performativity—its power to establish what qualifies as 'being'—works not only through reiteration, but through exclusion as well" [8]. While there are differences among subjects, data and researchers, the Semantic Web provides a way of challenging the hierarchical structure of the researcher and the researched, or audience and performer. Goffman provides a method for removing (or at least making visible and by default changing) the hierarchical structure and applying simultaneous contexts that are more reflective of the Semantic Web. For example, in one performance, a researcher is exposed to the data – its proceedings, its resource, etc. Simultaneously, the researcher can apply his or her intended performance onto the data; thus, the multiple effects that researchers may have on their research subjects surface. This act, in which the back region enters into the front region and thus modifies its performances as a result of this intrusion, has been posited by Goffman will result in "performers [who] will find themselves temporarily torn between two possible realities..." [12]. In this case, it will result in both the researchers and the researched playing simultaneous performative roles. However, the Semantic Web will allow for limitless realities as Semantic Web tools continue to be develop.

Another dividing line that may make regions distinct in the Semantic Web is the technological boundary that separates researcher and subject. While the encoded actions, as per the Semantic Web are computerized symbols, and the researcher works with those computerized symbols, there is nonetheless a dividing line between what the Semantic Web is capable of revealing and what, if anything, the researcher is able to view. This perspective is in keeping with a social constructivist theoretical approach that views regions and borders, and the separations between each, as human performative acts, similar to the coded (and uncoded) practices of the Semantic Web. Goffman wrote, "the line dividing the front and back regions is illustrated everywhere in society" as between users and researchers [12]. We believe that all of these experiences can (and should) be taken as individual performances. However, it is important to reiterate our reference to the performance of Semantic Web as a function of big data. Thus, we emphasize the many performances alongside one another, in conjunction with each other, and, at times, as a result of evolving interactions.

The researcher is often compelled to strategize masses of users, working collectively toward a goal. If we begin to apply performative aspects to the Semantic Web, we can ask: how might this change those who are researched or those who might perform out of character? Or more specifically, shield themselves from "the deterministic demands that surround them" [12]. But, perhaps even more encompassing, we must explore the following question: *how should researchers and their subjects, given the availability of data and how that data is "read" or interpreted, come to understand this data?* Butler reminds us to be cognizant of the "materialities" of conditions that enable and restrict categories that are embedded with history and historicity, each constructed by boundary lines, and thus enabling and constructing "both persistent and contested regions" [8].

Burke provides one specific way of conceptualizing the Semantic Web that we find useful, namely through his concept of terministic screens, through which ideas or concepts are highlighted, diffused, and directed [7]. Terministic screens operate as a means of understanding how something is understood and contextualized beyond merely observations. Terministic screens operate as perceptual devices that focus interpretation in keeping with the nature of whatever is being viewed through a terministic screen [7]. Terministic screens "...affect the nature of our observations, in the sense that the terms direct the attention to one field rather than to another" [7]. However, the Semantic Web provides a multitude of terministic screens that direct attention towards the many implicit and explicit ways in which things (in this case data) are linked. For Burke, the notion of multiple terministic screens was one to be embraced for each instance lends value to our understanding of nature.

THEORY IN PRACTICE

As we've highlighted, the explosion of available data continues to problematize what we study and how we study it. Thus far, we have called into review new technologies like Linked Open Data and the Semantic Web, which we suggest begin to temper concerns related to the encroachment of quantitative-only studies into the social sciences. We also theorized these new instruments to complicate our own understanding of the Semantic Web as it relates to critical theory in the social sciences. The remainder of the paper moves past explanation and theory and into implementation of Linked Open Data and semantic technologies. We take this opportunity to put into practice what the critical theory promises to theorize about these technologies and whether the technologies embody and test aspects of performance theory [17].

Population Science Grid (PopSci Grid)

In 2011, the Tetherless World leveraged a semantically enabled approach to integrate, visualize, and explore health data. Similar to the LOGD portal, one of the primary goals was to allow policy makers the opportunity to explore potential correlations between health-related policies and behavior change. In addition, the health data portal demonstrated the value of linking open data and semantic technologies for exploration of data by researchers and consumer audiences [24].

In collaboration with the Division of Cancer Control and Population Science at the National Cancer Institute (NCI), associated contractors, and university researchers from Rensselaer Polytechnic Institute (RPI) and Northwestern University, the team developed the Community Health Portal called PopSciGrid. The initial domain centered on tobacco policies, smoking prevalence, and related demographics [24]. Data was gathered from the ImpacTeen.org, the National Health Interview Survey (NHIS), the Health Information National Trends Survey (HINTS), State Level Tobacco Control Policy and Prevalence Database, and the National Cancer Institute etc.

Drawing from our discussion of Goffman, we examine how and if this semantically-enabled portal sets boundaries, while allowing for movement between regions - exploiting the performances of both the researcher and the subject. We begin with the act of converting the data into RDF, which is automated by a semantic-enabled tool developed by the TWC. This performance sets the separate regions by construct of the URI. We consider how once heterogenous pieces of data (which were implicitly hidden in the backstage behind incongruent datasets) are now explicitly revealed (moving into the front-stage) towards the researcher, politician, citizen, and so forth. Additionally, given the variety of data sources (e.g. NCI, ImpacTeen.org, NHIS, etc.) and agents involved in this collaborative effort, provenance (or histories of data) again permits for the movement between the back region to the front region. For example, a NCI agent may see the history of changes or additions made to data by a TWC agent. Thus, we suggest that the portal manifests to some degree ideas like a “region of boundaries”.

However, while PopSciGrid offers an excellent example of the power of semantic technologies, linked data and the web’s resources, its statistical based results reveal only part of the answer. Turning again to our discussion of Goffman, we are able to exploit shortfalls within the PopSciGrid performance. For example, while collaborators are empowered by the movement of data from the back region, the performance is limited by its exclusion of human encoded practices. Changes to the data are captured in provenance, but what of the motivations for the change? In this instance, we recognize the potential shortfalls of the selected social theory. While a performance on the data-level is important to consider, it only scratches the surface on whether semantic technologies in fact facilitate a better understanding of social behavior or society.

CONCLUSION

It is indisputable that the Web has transformed the way we interact as a society and as a culture. It is also without doubt that economic, political, social and cultural forces will continue to affect the development of this new network. As researchers, it is in then in our best interest to review how we effectively tackle new problems that may emerge in this space. Already, we see the merging of disciplines where the mixture of quantitative and qualitative data results is far more robust. In this hyper-connected world, it seems less and less likely for a researcher to remain within the boundary of a single field. Collaboration has certainly become part of our everyday lexicon. Yet, we argue that in order to construct a more complete picture of contemporary human behavior, researchers must continue to evolve its tools and methods.

Tools, like linked open data and semantic technologies, present an opportunity to capture data and all acts of interpretation of the data. Concerns related to authenticity and authority are thus explicitly encoded. The goal is to make the implicit, explicit in the data. Moreover, bridging datasets into a more graph-like structure broadens the scope of the data. To further illustrate, we presented both Goffman’s and Burke’s social theories of performance and the terministic screens as they relate to the Semantic Web. Our goal was to demonstrate that through these theoretical frameworks, researchers may begin to re-conceptualize how big data can be collected, analyzed, shared, and synthesized. In Pickering’s words, “in the performative idiom, one sees that scientific knowledge attaches, visibly, not to the ‘world itself’, but to specific mechanic fields, distributions of specifically tuned machines and instruments” [30].

More importantly, in order to move beyond just multi-disciplinary collaboration, researchers must consider the cultural and conceptual differences across sciences so that more complex problems maybe addressed [4]. Here we caution that being more complete does not mean absolute, perfect, or consistent. Just as the World Wide Web is a digital space for exploration of ideas, disagreements, different truths and facts and ideologies, so too is the Semantic Web.

Foucauldian tradition in the discussion of "truth" is to be understood as a system of ordered procedures for the production, regulation, distribution, circulation, and operation of statements. "'Truth' is linked in a circular relation with systems of power that produce and sustain it, and to effects of power which it induces and which extend it..." [11]. When applied to understanding the Semantic Web, we highlight the technology's allowance of different viewpoints, ideas, or truth at the data level making explicit where exactly these disagreements take place. Statements on the Semantic Web are only true for those who make and use them. There is no notion of absolute or objective truth, just as we find in the Web. Having provenance (histories) of these data be linked to the data, allows us to know who made these statements, what authority they have to make such statements, and whether we should trust them or not. Moreover, the Semantic Web, in its current construction, exists within these ordered systems (or algorithms) and have been written about as having subversive and/or hegemonic properties, namely its ability for surveillance. However, this is beyond the scope of this paper.

The Semantic Web also operates under the Foucauldian idea of *parrhesia* in which what is produced is a "complete and exact account of what he [a user or a research subject] has in mind" [11]. The recordings, from minutiae to large data sets, become a representation of self. However, this is also subject to debate, especially if we see the Semantic Web in producing multiple truths. "The speaker makes it manifestly clear and obvious that what he says is his own opinion" [11]. And while it can be perhaps argued that a research subject is not necessarily producing an act of truth, within the confines of Foucauldian parrhesia, perhaps the literal recording of data sets, like the recording or transcription of a public event, becomes the most accurate truth currently available. Semantic Web data does occur (the actual data itself, not the decoding of any data) outside the traditional practice of rhetoric. More importantly, the Semantic Web, by design takes an open world assumption, which means the absences of knowledge or a statement does not mean that statement to be false. This shows the idea of incomplete information in a system. This system contains many different aspects that are evolving alongside semantic web tools; and, will ultimately reveal data as much as it may possibly obfuscate data and interpretations of that data. Nonetheless, just because we don't have knowledge that something exists does not mean it doesn't exist at all.

ACKNOWLEDGMENTS

We thank all the Tetherless World Constellation lab students and faculty, who provided helpful comments on previous versions of this document.

REFERENCES

1. Anderson, C., "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete". *Wired* Magazine. (2008). http://www.wired.com/science/discoveries/magazine/16-07/pb_theory/.
2. Arroyo, L., "Social Web: How can we STUDY the Social Web?: The Web Science." Lecture at the VU University Amsterdam. Given March 11, 2013. Slideshare. http://www.slideshare.net/laroyo/lecture-6-how-do-we-study-the-social-web-2013
3. Berners-Lee, T., Hendler, J., & Lassila, O., The Semantic Web. *Scientific American*. (2001), 29-37. http://www.scientificamerican.com/article.cfm?id=the-semantic-web.
4. Bigras, E. "Innovation in Asthma Research: Using Ethnography to Study a Global Health Problem". Ethnography Matters. (2012). http://ethnographymatters.net/2012/10/27/the-asthma-files/
5. boyd, d. & Crawford, K., *Six Provocations for Big Data*. "A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society". (2011). http://ssrn.com/abstract=1926431.
6. Blumer, H., *Symbolic Interactionism: Perspective and Method*. Prentice Hall, NJ, USA, 1996.
7. Burke, K., *Language as Symbolic Action: Essays on Life, Literature, and Method*. University of California Press, Berkeley, CA, USA, 1966.
8. Butler, J., *Bodies That Matter: On the Discursive Limits of "Sex"*. New York: Routledge, 1993. Print.
9. Deacon, D., Pickering, M., Golding, P., & Murdock, G., *Researching communications: a practical guide to methods in media and cultural analysis*. (2nd ed.). Bloomsbury, USA, 2007.
10. Ding, L., Lebo, T., Erickson, J. S., DiFranzo, D., Williams, G.T., Li, X., Michaelis, J., . . .Hendler, J. WC LOGD: A Portal for Linked Open Government Data Ecosystems. *Journal of Web Semantics*. (2011). http://logd.tw.rpi.edu/2011/logd-jws2011swc-v1.98.pdf.

11. Fontana, A. & Pasquale, P. (Interviewers) & Foucault, M. (Interviewee)., "Truth and Power." Published in *Microfisica del potere: interventi politici,* 1977, 3-28 in *The Essential Foucault*. Ed. Paul Rabinow & Nikolas Rose. Norton, NY, 2003.
12. Goffman, E., *The Presentation of Self in Everyday Life*. Anchor Books, New York, NY, USA, 1959.
13. Goffman, E., *Frame Analysis: An Essay on the Organization of Experience*. Cambridge, Mass: Harvard University Press, 1974. Print.
14. Grice, P., *Studies in the Way of Words*. Harvard University Press, Cambridge, MA, USA, 1989.
15. Hall, W., "The Ever Evolving Web: the Power of Networks". *International Journal of Communication*, (2011), 5, 651-664. http://ijoc.org/ojs/index.php/ijoc/article/download/1120/548. 659.
16. Hendler, J., *Broad Data*. [PowerPoint Slides], (2012), http://www.slideshare.net/jahendler/broad-data.
17. Landow, G., "Hypertext and Critical Theory." *Reading Digital Culture*. Ed. David Trend. Oxford and Malden: Blackwell, (2001), 98-108.
18. Laplace, P.S., *A Philosophical Essay on Probabilities*. Chapman & Hall. New York, NY, USA, (1902), 177.
19. Latour, B., 'Tarde's idea of quantification', in The Social After Gabriel Tarde: Debates and Assessments, ed M. Candea, London: Routledge, (2009). 145-162. www.brunolatour.fr/articles/article/116-TARDE-CANDEA.pdf
20. Lehrer, J. *Proust was a Neuroscientist*. First Mariner Books. New York, NY, USA, 2008.
21. Li, X., Lebo, T., & McGuinness, D.L. Provenance-based strategies to develop trust in semantic web applications, in: The Third International Provenance and Annotation Workshop, 2010, 182–197.
22. Manovich, L., "Trending: The Promises and the Challenges of Big Social Data". Debates inthe Digital Humanities, ed M.K.Gold. The University of Minnesota Press, Minneapolis, MN, USA, (2011). http://www.manovich.net/DOCS/Manovich_trending_paper.pdf
23. McGuinness, D.L., Furtado, V., daSilva, P.P., Ding, L., Glass, A., & Chang, C., Explaining semantic web applications, in: *Semantic Web Engineering in the Knowledge Society, Information Science Reference*, (2008), 1–24.
24. McGuinness, D.L., Shaikh, A. R., Moser, R., Hesse, B., Morgan, G.,…Devlin, H., "A Semantically-enabled Community Health Portal for Cancer Prevention and Control." Presented at the Web Science Conference 2011. (2011). http://tw.rpi.edu/media/2011/07/14/8c95/mcguinness2011popscigrid-paper.pdf
25. McNely, B., "Big Data, Situated People: Humane Approaches to Communication Design". *Communication Design Quarterly Review* (2011), http://sigdoc.acm.org/wp-content/uploads/2012/09/CDQR_1-1_Fall2012.pdf.
26. Meyer, E. & Dutton, W.H., "Top−Down e−Infrastructure Meets Bottom−Up Research Innovation: The Social Shaping of e−Research. *Prometheus*. (2009), 27(3), 239-250. doi: 10.1080/08109020903127810
27. Michaelis, J., & McGuinness, D.L., "Towards provenance aware comment tracking for web applications", in: The Third International Provenance and Annotation Workshop (2010), 265–273.
28. Murphy, R., "Use Case 8 - Linked Open Government. W3C" eGov Wiki. (2008). http://www.w3.org/egov/wiki/Use_Case_8_-_Linked_Open_Government
29. New York Public Colleges. *College Completion*. Retrieved from The Chronicles Of Higher Education Website. (n.d.). http://collegecompletion.chronicle.com/state/#state=ny&sector=public_four
30. Pickering, Andy. "After Representation: Science Studies in the Performative Idiom." *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1994 (1994): 413–419. Web. 25 Mar. 2013.
31. Rabinow, P. & Rose, N. *The Essential Foucault*. Ed. "Truth and Power." (n.d.) http://foucault.info/documents/parrhesia/foucault.DT1.wordParrhesia.en.html
32. Searle, J. R. *Speech Acts: An essay in the Philosophy of Language*. Cambridge at the University Press, London, 1969.
33. Shadbolt, N., Berners-Lee, T., & Hall, W., "The Semantic Web Revisited". *IEEE Intelligent Systems*. (2006), 21(3), 96-101. http://eprints.ecs.soton.ac.uk/12614/01/Semantic_Web_Revisted.pdf
34. Stone, B. E., "Subjectivity and truth" in *Michel Foucault: Key Concepts*. Ed. Dianna Taylor. Acumen, Durham, UK, 2011.
35. Wilfford, C. & Henry, C., 'One Culture': Computationally Intensive Research in the Humanities and Social Sciences: A Report on the Experiences of First Respondents to the Digging Into Data Challenge.

Council on Library and Information Resources.151. (2012). http://www.clir.org/pubs/reports/pub151

36. Williams, D., Contractor, N., Poole, M. S., Srivastava, J., & Cai, D., "The Virtual Worlds Exploratorium: Using large-scale data and computational techniques for communication research." *Communication Methods and Measures,* (2011), *5*(2), 163-180. doi: 10.1080/19312458.2011.568373

37. Woodhouse, E.J., "Information Technology, Muddling through and the potential Intelligence of democracy." [Draft]. (2000) http://homepages.rpi.edu/~woodhe/pdfs/midwest.pdf

38. World Wide Web Consortium. "Semantic Web". (n.d.) http://www.w3.org/standards/semanticweb/

39. van Ossenbruggen, J. R., Hardman, L., & Rutledge, L., "Hypermedia And The Semantic Web: A Research Agenda." *Journal of Digital Information*. (2002), *3*(1). http://jodi.ecs.soton.ac.uk/Articles/v03/i01/VanOssenbruggen/