HISTORICAL INFRASTRUCTURES FOR WEB ARCHIVING:

ANNOTATION OF EPHEMERAL COLLECTIONS

FOR RESEARCHERS AND CULTURAL HERITAGE INSTITUTIONS

Meghan Dougherty
SUNY, Institute of Technology
meghan.dougherty@sunyit.edu

Charles van den Heuvel
Virtual Knowledge Studio
charles.vandenheuvel@vks.knaw.nl

April 2009

HISTORICAL INFRASTRUCTURES FOR WEB ARCHIVING

Abstract

The World Wide Web is becoming a source of information for researchers, who are more aware of the possibilities for collections of Internet content as resources. Some have begun creating archives of web content for social science and humanities research.

However, there is a growing gulf between policies shared between global and national institutions creating web archives and the practices of researchers making use of the archives. Each set of stakeholders finds the others' web archiving contributions less applicable to their own field. Institutions find the contributions of researchers to be too narrow to meet the needs of the institution's audience, and researchers find the contributions of institutions to be too broad to meet the needs of their research methods. Resources are extended to advance both institutional and researcher tools, but the gulf between the two is persistent.

Institutions generally produce web archives that are broad in scope but with limited access and enrichment tools. The design of common access interfaces, such as the Internet Archive's Wayback Machine, limit access points to archives to only URL and date. This narrow access limits the ways in which web archives can be valuable for exploring research questions in the humanities and social sciences. Individual scholars, in catering to their own disciplinary and methodological needs, produce web archives that are narrow in scope, and whose access and enrichment tools are personalized to work within the boundaries of the project for which the web archive was built.

There is no way to explore a subset of an archive by topic, event, or idea. The current search paradigm in web archiving access tools is built primarily on retrieval, not discovery. We suggest that there is a need for extensible tools to enhance access to and enrichment of web archives to make them more readily reusable and so, more valuable for both institutions and researchers, and that annotation activities can serve as one potential guide for development of such tools to bridge the divide.

The contextual knowledge production evolving from annotation not only adds value to web archives by providing one solution to the problem of limited resources for generating metadata in web archives; it also forms part of our collective memory and needs to be preserved together with the original content. In the 19th and 20th centuries documentalists, such as Paul Otlet (1868-1944) began exploring methods to order, access, and annotate ephemeral, dynamic material for research. Otlet developed a documentation system in which bibliographical material describing content transmitted by all sorts of media (radio, film, gramophone and television) was stored together with various forms of annotations, ranging from updates to expressions of opinion. It imagined researchers working together on a global level to create and to enrich collective memory. We claim that these pre-web annotation initiatives are also of interest for future strategies to access and preserve more dynamic and ephemeral forms of digital cultural heritage, such as web archiving.

Web Archives for Research

The web is a source of information for humanities and social science scholars who are increasingly aware of the possibilities for using collections of web content as research resources. The web is a resource to locate relevant information, it is an objects of analysis, and a resource for studying the evolution of scholarly practice. For these scholars, the Internet has become a place for organizing information and building new research tools. Given this shift in attention toward online culture — both studying it, and incorporating it into research practices — digital cultural heritage resources, such as web archives, are becoming fundamental assets to humanities and social sciences researchers. Web archiving is growing in its own right, and this growth and the value it can offer to humanities and social sciences, depends on steady development of tools, standards, policies, and services upon which researchers using digital cultural heritage in their research can rely.

Web archiving is in a state of flux where boundaries around traditional roles of researchers and stewards are blurring. This blur is resulting in much experimentation with different practices guided by different motivating principles. Some social science and humanities researchers have begun creating their own research-driven archives of web content (sometimes partnered with larger information institutions, such as the Internet Archive, the Library of Congress, or national libraries throughout Europe), and encouraging others to do the same.[1] Others in information design have partnered with the same information institutions to create web archives and interfaces that have a broader appeal.[2]

Web archiving experimental trials conducted by a social scientist, a linguist, or an historian will inevitably result in archives that differ from each other, and that are fundamentally different from an archive resulting from the experiments of a librarian. Each practitioner is motivated by a different mission, be it institutional, methodological, ontological, or epistemological. Diverse approaches to web archiving are resulting from these experiments and is increasingly leading to

conversation and collaboration across fields to develop inclusive practices. The positive outcomes of experimentation in web archiving should be sustained and used to promote new insights.

This varied experimentation has led people in different fields to explore possibilities for ongoing interpretation within archived collections, the accumulation of communicative traces, reinterpreting and regrouping - cognitive wandering[3] - to make meaning of archived artifacts. The networked origins of these artifacts suggest a social medium for performing the work of interpretation and knowledge production. One possible solution proposed in this paper draws on annotation and collective memory to redefine preservation and interpretation for web archiving. This redefinition opens possibilities for inclusion and representation in stewardship and analysis of digital cultural heritage to mirror the openness and cooperativeness of cybercultural media environments.

Digital artifacts often lack history, texture and depth. Compared to a used book or a public bulletin board, most digital artifacts do not accumulate rich marks made by previous users. These marks serve as pointers to help us interpret information, and preserve evolving cultural heritage. Because we often fail to see the evolutionary history of cybercultural artifacts – the rich marks are invisible and occur only within the individual user's experience of web artifacts or the researcher's project-specific analysis – we disregard artifacts in cybercultural history in general. The evolution of ephemeral artifacts on the web becomes the purview of digital scholarship and a side note in history, and without rich access points for researchers even this scholarly history remains difficult to explore. Traces left by previous users alter the history of real-world and cybercultural objects, either by changing the object itself or changing the use or meaning of the object over time. These annotations, when considered significant for the preservation of cultural heritage, can provide the history, texture and depth that cybercultural artifacts currently lack, and develop rich access points for researchers. Annotations kept as descriptors of artifacts could become the most important criterion for which knowledge to preserve for the future. Scholars would not only navigate, but also discover new information in this continuously updated and enriched collective memory network of humanity's knowledge. But, before the value of annotations can be realized, practices among different parties should be considered.

Divergent Paths for Practices

There is a growing gulf in web archiving between the researchers who want to use web artifacts to study in their field and the information professional who serve information needs. The divergent paths on either side of this gulf can be seen on different levels, be it the differences between global and national institutions, or individual researchers and information professionals. While the contributions made by these stakeholders all advance web archiving as a whole, their practices reveal a gulf between institutional and personalized perspectives. Each set of stakeholders finds the others' contributions to web archiving less and less applicable to their own field. Large institutions find the contributions of researchers and small targeted collections to be too narrow to meet the diverse needs of the large institution's audience. Researchers find the contributions of institutions to be too broad to meet the focused needs of their research. Resources are rightly extended to advance the institutional and personalized tools in web archiving to bolster cyberinfrastructure for humanities and social sciences, but the gulf between the two is persistent. The nascent state of web archiving practices provides an opportunity to address that gulf.

Despite increased attention to web archiving in e-heritage and e-research, there are two problems that remain obstacles for the development and subsequent use of web archives — access and enrichment. These obstacles have been largely side-stepped in favor of meeting other technological needs. The developments that have addressed these obstacles have been made in small-scale individual projects. Large-scale reusable collections have begun to create tools to address these challenges, but they do not meet the needs of scholars who want to use the resources.

Institutions generally produce web archives that are broad in scope, but have limited access points, and few, if any, enrichment tools. The design of commonly used access interfaces, such as the Internet Archive's Wayback Machine, limit access points to web archives to only URL, date, and full text search. This kind of access is valuable for only a handful of approaches to social

science and humanities research - it limits the types of research questions that can be asked of the materials. It does not only limit the number of questions that can be asked, but also limits the approaches that can drive long-standing research paradigms. These limited access points favor questions of content and producer. It makes other types of questions difficult to ask, and other methods for collecting data difficult to execute. This narrow access limits the ways in which web archives can be valuable for exploring research questions in the humanities and social sciences, and if not limiting, then certainly influential in the styles and methods of research conducted in web historiography and other methods employing web archives.

On the other hand, individual scholars, in catering to their own disciplinary and methodological need, produce web archives that are narrow in scope, and whose scope and content, and whose access and enrichment tools are personalized to work only within the boundaries of the project for which the web archive was built. It is often the case that these specialized archives are largely valuable only in reference to the immediately surrounding study. These types of specialized archives may be valuable in reference to each other, as comparative studies or small samples, but systems to draw disparate archives together, or to compare artifacts or data points across archives do not yet exist.

There is no way, yet, to explore a subset of an archive by topic, event, or idea. The current search paradigm in web archiving access tools is built primarily on the basis of retrieval, not discovery. We suggest that there is a need for extensible tools to enhance access to and enrichment of web archives to make them more readily reusable and so, more valuable for both institutions and researchers, and that annotation activities can serve as one potential guide for development of such tools to bridge the divide.

There exists an opportunity to examine approaches in e-heritage, and shift them to be more inclusive in order to add value for researchers in the humanities and social sciences. This opportunity stems from the current malleable state of practices in web archiving. Practices within library and information science have come to dominate because the practices built in these fields are technologically advanced and ready to handle the content delivery systems required in web archives

- they offer a system that can serve as an offline analogy from which we can build web archiving. The consequence of this is that library and information science tropes have permeated the development of web archives regardless of why those archives were developed or how they might be used. This has set up a point of contention between librarians and information architects who would like to build widely valuable and accessible collections, and humanities and social science researchers who would like to form web archiving as a method for developing digital cultural heritage or web historiography. The two perspectives are not diametrically opposed, but there are certainly points of contention that are derived from differently held philosophical undercurrents that motivate each field.

Genuine progress is being made from both sides of the divide. Large libraries and archives continue with their efforts to build large multi-purpose web archives that further institutional mission, and researchers — either on their own, or partnering with archivists — develop their own archives for use in their research. But, national archives cannot justify the use of resources to build the directed, project-specific archives that researchers build when given the opportunity. And, researchers cannot yet find value for their focused collections in the large multi-purpose archives being built by institutions. There is a call for exploration in the development of cyberinfrastructure for e-heritage and e-research[4], and resources are being allocated to further this, especially for humanities and socials science, but efforts continue to result in a disconnect between institutional cyberinfrastructure development and the needs of the humanities and social science researchers they aim to serve.

Of the multiplicity of avenues that should be explored and multiplicity of tools that should be developed as solutions to this problem, we will address one avenue – shared annotation. Access and enrichment can be aided by annotation tools as one simple solution. The contextual knowledge production evolving from annotation not only adds value to web archives by providing one solution to the problem of limited resources for generating metadata in web archives; it also forms part of our collective memory and can be preserved together with the original content.

Calling on Social Solutions for Guidance

What would happen if we applied what we know of the semantic web, search strategies, swarming, and linking theories to museum and library activities of cataloging and exhibition, museological, and scholarly ambitions of post-structural knowledge production? Would we no longer consider that the meaningfulness in cultural artifacts requires an authoritative interpretation? Might we find value in multiple and socially-generated meanings about cultural heritage? For researchers, we may find that sifting through artifacts in web archives, searching for relevant objects of analysis, seeking out appropriately sized samples may evolve into a conversation among a range of scholars, and may open the field of research approaches beyond existing web archive search paradigms.

There are alternatives to authoritative systems that derive their management from strict process, workflow, security and control. These alternatives employ user-driven meaning making as part of the process of creating accessibility. These notions are not new; there are historical processes in information management that challenge the hierarchical information management that cannot include the deep contextual and cultural usage meanings that might easily place one object in multiple categories. The restrictions that arise from employing authoritative management of knowledge can be avoided with the participatory, inclusive, social and representative knowledge ecology that is fostered by information networks.

Such systems that incorporate varying levels of user-supplied annotation, are valuable because they bolster the user - researcher or otherwise - who feels she has contributed to a community.[5] Annotation, including note-taking and tagging, does more than just provide a sort of bookmark or placeholder for an archive user. This type of participatory media ecology as applied to archive access could account for changes in interpretation over time, negotiation of interpretation, and possibilities for including new interpretations.

Process, workflow, security, and control issues that pervade knowledge management are still important issues. But, some organizations are branching out from this paradigm by incorporating more inclusive and participatory tools.[6]

These experimental tools recognize that inclusive and social interfaces to digital artifacts frame archive objects as containing cultural artifacts that are polysemic entities. This is no different when applying these concepts to web archives as well. The meanings of narratives and classificatory systems used to describe web archive objects are products of disciplinary perspective. This is noticeably present in narrowly-coded, scholarly, project-specific archives, and noticeable absent in the meagerly documented, large-scale, institutional archives. These opinions are mediated through access tools and installations as interfaces or representations that produce knowledge for various viewers in various contexts. In these various contexts, a diverse range of actors engage in the cycle of knowledge production. By extending redefined authority and polysemy to processes of interpretation, representation, and preservation, artifacts of networked culture can be preserved as history-rich objects showing accumulated user interaction, and interpretation.

Historical Infrastructures for Social Tagging and Collective Memory

As noted earlier, we contend that the contextual knowledge production evolving from annotation not only adds value to web archives by providing one solution to the problem of limited resources for generating metadata in web archives; it also forms part of our collective memory and needs to be preserved together with the original content. In the 19th and 20th centuries, documentalists such as Paul Otlet (1868-1944) began exploring methods to order, access, and annotate ephemeral, dynamic material for research.[7] Otlet developed a documentation system in which bibliographic material describing content transmitted by all sorts of media (radio, film, gramophone and television) was stored together with various forms of annotations, ranging from updates to expressions of opinion. It imagined researchers working together on a global level to create and to enrich collective memory. We claim that these pre-web annotation initiative may be of interest for future strategies to access and preserve more dynamic and ephemeral forms of digital cultural heritage, such as web archiving.

Dynamic Documenting for Research: An Early 20th Century Example

In 1913 Paul Otlet advocated to let researchers united in networks and individual scholars to work together on what he called, the Livre *universal de la Science.* This universal book of science Otlet envisioned as " an unlimited work, always up-to-date, constantly growing, concentrating, absorbing, synthesizing, systematizing every intellectual product from the moment it is born".[8] At first sight, it seems an early form of Wikipedia, in an open, distributed network, similar to the Web. However, the positivist Otlet believed too much in scientific objectivity to leave knowledge creation to the wisdom of the crowd. It was the critical mass of the crowd Otlet was after, but the implementation of such collective enterprises was streamlined in much detail.[9] The controlling element that hold this knowledge production processes together was the Universal Decimal Classification System (UDC), that Otlet developed on the basis of Mevil Dewey's classification system. For our comparison with web archiving and collaboratories in research it is important to note that the control system of the UDC had features in common with hypertext and shared databases.[10]

Different from purely topical classification schemes, the UDC did not just order subjects or topic in classes by numeric codes, but also allowed for linking to additional facets, such as place, language, physical characteristics via its auxiliary tables of connector signs. It is a system of related parts that by numeric codes and connectors, such as "+, / and :" provided: "the links, the genealogy even, of ideas and objects, their relationships of dependence and subordination, of similarity and difference".[11] On a technical level the classification system made it possible to link annotations to specific documents, or parts of (interrelated) documents around a classification number. The linkage characteristics of the UDC would not only allow connecting various classification systems (see figure 1), but also creating a space of contributors around documents.

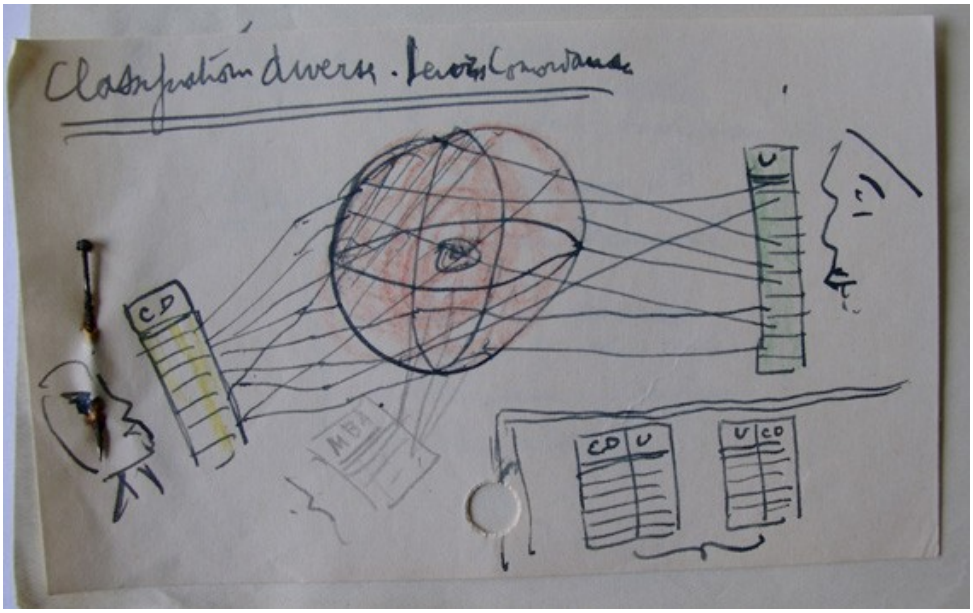

Figure 1. Otlet linking various classification systems – EUM- 1-4 – Mons, Mundaneum ©

The latter could revise document in the form of annotations, ranging from additions to various points of view (see, figure 2).

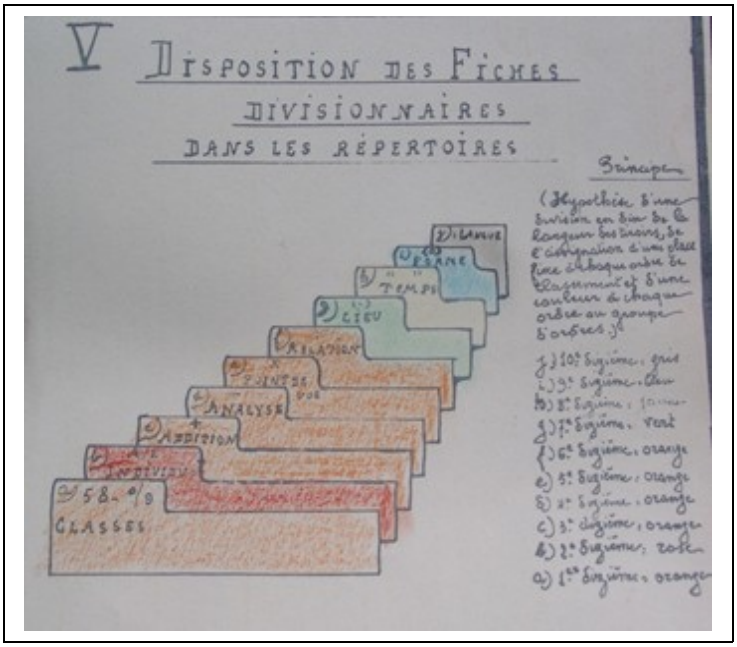

Figure 2. Otlet – EUM – Affiche Table de classification – 58 botanique (detail) – Mons Mundaneum ©

Card box with colored tabs for annotations in relation to knowledge class. Knowledge class (orange) in relation to name of the annotator (red) whose input is classified as addition, analysis, point of view or relation to other subjects (orange tabs), and in relation to space (green), time (beige), form or medium (blue), and language (gray).

These links were made manually, but Otlet also studied mechanical ways to create a system of ideas, a mechanical brain. This mechanical brain should not just serve the intellectual work of collaborating scholars, but in principal everyone: “Like the technical machine allows not qualified workers to make perfect products, the intellectual machine does not require a specific education of the one that uses it”.[12]

This statement is interesting for our exploration of the preservation of website archives with annotations for collective memory and suggests the need for a closer look at the infrastructure that Otlet proposes for personal and collective data enrichment. Moreover it brings us to the ongoing debate on the authority of experts and lay-experts in Web 2.0.

Past and Future Tagging

For Otlet the process of documentation did not only involve the creation of a knowledge system; it was a social system aimed at creating a better society. “One can imagine a social state that makes progress in its whole by an instrumentation based on very high levels of abstraction that would be made available to everyone”.[13] But where stands the individual in all of this?

Otlet’s concept of personal knowledge organization is strongly related to, is actually a microcosm of his universal classification system. Otlet recognized the value of extracting personal notes from documents: “Preserved, classified, revised, continuously enriched with other notes derived from other sources, they could become a real book: a particular book for each person of which one could say: ‘My Book’, ‘My Encyclopedia’[…]an artificial memory of everything one desires to recall”.[14] He also believed in the value of preserving these annotations for collective memory. By classifying and storing the notes together with bibliographical descriptions of documents, “One could avoid new transcriptions often subject to errors, keep up with facts and ideas annotated at various moments. The confirmations by others that may also express different aspects of the same thing”.[15]

Alex Wright states that Otlet’s vision allows marrying top-down classification systems such as the UDC with socially-constructed information spaces, such as MySpace.com, Flickr.com

and del.icio.us with their own folksonomies and tags.[16] The process seems at first sight indeed similar to Wikipedia in which the involvement of more people adding and editing certain lemmas in the end to the improvement of those lemmas in particular and the digital encyclopedia in general. However, for annotating documents Otlet had eminent scholars in mind who, regulated by protocols for intellectual work, would further develop the Universal Network of Documentation and the UDC to reach ever higher levels of scientific objectivity.[17]

Although Wright is correct in a technical sense, it would be wrong to read Otlet's Universal Network of Documentation simply as a wiki or his "personal classifications" as folksonomy. For Otlet the producer of the knowledge is foremost an outsider of the system whose contribution would only be recognized after a long process of editing by what we would call nowadays domain experts.[18] Compared to Wikipedia, Otlet's knowledge system and collaboratory is more top down, but at the same time also more transparent. Edits and annotations do not merge directly with the information, but remain visible in an ordered way, describing the provenance and intention of the proposed data enrichment of the Universal Documentation Network. As such we might consider it as part of the proposed infrastructure to preserve website archives and annotations as a collective memory for research.

We have argued that large infrastructures that preserve web archives together with other digital repositories, such as the Internet Archive, do not have the right interfaces and annotation systems yet, to exploit them fully for research. Existing systems such Hanzo Archives or iBreadCrumbs allow researchers to annotate website archives, but annotations are not kept together with these archives to create a collective memory. Following Wright, we claimed that Otlet's system of auxiliary tables in a technical sense allowed keeping together dynamic annotations with changing content. At the same time we have challenged Wright's idea to read Otlet's Universal Documentation Network as a social space, such as Wikipedia. The nature of Otlet's knowledge network, is too hierarchical for that. However, its transparency might be usefully investigated in discussions of the role of authority in distributed authorship.

The attempt by Otlet to uphold scientific authority by designing protocols for scholarly collaboration and developing classifications or typologies of annotations comes close to recent attempts to differentiate in forms of expertise in Web 2.0. The hyperlink structure of Web 2.0 in principal allows providing context to digital resources. Moreover, it might allow for the necessary critical mass, that in a traditional way such as in the small, heterogeneous datasets used in the humanities, is not always possible. Chuck Zerby observes in his description of the process towards what he calls the virtual footnote that: "As the footnote reconfigures itself for the digital world, opportunity and danger are waiting side by side for it".[19] Jenny Lyn Bader published in 2000 an article in The New York Times with the headline: "Forget Footnotes. Hyperlink. Old Media, Meet New Media" in which she claimed that after the eviction of the footnote by book publishers, they would find a new home in the hyper-link construction of the World Wide Web. "Indeed the Web has not only revived the footnote, it has spawned a cross-referencing craze that renders the formerly complete media event into a […] wallflower waiting to be courted by the next available annotator".[20] However, it is fair to say that many researchers ,especially in the humanities and social sciences do not trust tags in the same way, as for the traditional footnote to put their scientific argument into context and therefore are hesitant to become annotators. In the ACLS report Our Cultural Commonwealth on cyberinfrastucture in the humanities and social sciences this hesitation was labeled as "conservative culture in scholarship".[21] However, it might be more fruitful to explore possible causes of conservatism in e-research.[22] Paul Wouters mentioned the mismatch between this cyberinfrastructure and the digital scholarly practices.[23] For annotation in e-research in the last years some infrastructures were explored to combine social tagging with expertise-based authority.

Larry Sanger, co-founder of Wikipedia, built a new user-generated encyclopedia, Citizendium, reviewed by domain experts. Software developers try integrating top down classification with bottom up tagging with a product called Facetag[24]; others give identity to links by visualizing their provenance in HarvANA[25]; others develop systems to assessing link value.[26]

The question whether the future lies in structuring information by experts or the collective wisdom of the crowd is ideological rather than technological. The infrastructures that we

envision to discover, preserve, navigate and enrich web archives as documents for research will shape that future. The way knowledge is organized, (re-)used and enriched by experts and non experts all over the world makes part of the cultural history of mankind. It requires an historical approach of knowledge organization and annotation in relation to practices of use.

Bibliography

American Council of Learned Societies, Commission on Cyberinfrastructure for the Humanities and Social Sciences. *Our Cultural Commonwealth: The Report of the American Council of Learned Societies' Commission on Cyberinfrastructure for Humanities and Social Sciences*. New York, NY: American Council of Learned Societies, 2006.

Anderson, Robert Charles, Barkley, Paul, Booth, Robert, Hilsclaw, Birdie, Velke, Robert, and Wylie, John Vincent. 2000. Gentech: Genealogical Data Model Phase 1: A Comprehensive Data Model for Genealogical Research and Analysis. http://xml.coverpages.org/genealogy.html#gentech (accessed April 9, 2009).

Bader, Jenny Lynn. "Forget Footnotes. Hyperlink. Old Media, Meet New Media." *New York Times*, July 16 2000.

Dougherty, Meghan. "Archiving the Web: Collection, Documentation, Display and Shifting Knowledge Production Paradigms." University of Washington, 2007.

Foot, Kirsten A, and Steven M Schneider. "Object-Oriented Web Historiography - Keynote Address." Paper presented at the Web_site Histories - Theories, Methods, Analysis, Aarhus, Denmark, October 14 2008.

Heuvel, C. van den (forthcoming). Web Archiving in research and historical global collaboratories. Interfaces and annotations to web archives in the humanities and social sciences seen from a historical perspective. Paper presented at Web_site Histories - Theories, Methods, Analysis 2008, October 14, Aarhus, Denmark. In N. Brügger ed. Web History, publication foreseen in the book series Digital Formation [ed. Steve Jones] by Peter Lang Publishing Group.

Hjörland, B., and Albrechtsen, H. "Toward a New Horizon in Information Science: Domain Analysis." *Journal of the American Society of Information Science* 46, no. 6 (1995): 400-25.

Hunter, J, I Khan, and A Gerber. "HarVana - Harvesting Community Tags to Enrich Collection Metadata." Paper presented at the Joint Conference on Digital Libraries, Pittsburgh, PA, June 16-20 2008.

Meyer, E, R Schroeder, and C Madsen. "The World Wide Web of Humanities: Archives for Researching the Web." Paper presented at Web_site Histories - Theories, Methods, Analysis 2008, October 14, Aarhus, Denmark. In N. Brügger ed. Web History, publication foreseen in the book series Digital Formation [ed. Steve Jones] by Peter Lang Publishing Group.

Otlet, Paul. "Le Livre Dans Les Sciences." Paper presented at the Conférence faite à la maison du Livre, Bruxelles, Musée du Livre, fasc 25-26 1913.

Otlet, Paul. *Traité De Documentation : Le Livre Sur Le Livre : Théorie Et Pratique*. Bruxelles: Editiones Mundaneum, Palais Mondial, 1934.

Otlet, Paul. *Monde, Essai D'universalisme : Connaissance Du Monde, Sentiment Du Monde, Action Organisée Et Plan Du Monde*. Bruxelles: Editiones Mundaneum/D.van Keerberghen&Fils, 1935.

Otlet, Paul. "Réseau Universel De Documentation: Documentatio Universalis (Rete)." *IID Communicationes* 4, no. 1 (1937): 13-16.

Quintarelli, Emanuele, Andrea Resmini, and Luca Rosati. "Facetag: Integrating Bottom-Up and Top-Down Classification in a Social Tagging System." Paper presented at the IA Summit, Las Vegas, NV, March 22-26, 2007.

Rayward, W B. “The International Exposition and the World Documentation Congress, Paris, 1937.” *Library Quarterly* 53, no. July (1983): 254-68.

Rayward, W B. “Visions of Xanadu: Paul Otlet (1868-1944) and Hypertext.” In *Historical Studies in Information Science*, edited by T. Bellardo Hahn and M. Buckland, 65-80. Medford, NJ: Information Today Inc, 1998.

Schneider, Steven M, Kirsten A Foot, and Paul Wouters. “Web Archiving in E-Research.” In *E-Research: Transformation in Scholarly Practice*, edited by Nick Jankowski, London: Routledge, 2009.

Wouters, Paul. “Forging New Academic Identities: Key Dilemmas and Challenges in Digital Scholarship.” Paper presented at the University of Washington Simpson Center, Seattle, WA, May 1, 2007.

Wright, Alex. *Glut: Mastering Information Through the Ages*. Washington, D.C.: Joseph Henry Press, 2007.

Zerby, Chuck. *The Devil's Details: A History of Footnotes*. New York, NY: Touchstone, 2003.

1 Kirsten A. Foot, and Steven M. Schneider, “Object-Oriented Web Historiography” Web_site Histories - Theories, Methods, Analysis Keynote Address, Aarhus, Denmark, October 14 2008.
2 Eric Meyer, et al., “The World Wide Web of Humanities: Archives for Researching the Web” (paper presented at Web_site Histories - Theories, Methods, Analysis, Aarhus, Denmark, October 14, 2008). In N. Brügger ed. Web History, publication foreseen in the book series Digital Formation [ed. Steve Jones] by Peter Lang Publishing Group.
3 Meghan Dougherty, “Archiving the Web: Collection, Documentation, Display and Shifting Knowledge Production Paradigms” (Ph.D. diss., University of Washington, 2007).
4 American Council of Learned Societies, Commission on Cyberinfrastructure for the Humanities and Social Sciences. *Our Cultural Commonwealth: The Report of the American Council of Learned Societies' Commission on Cyberinfrastructure for Humanities and Social Sciences*. New York, NY: American Council of Learned Societies, 2006.
5 Dougherty, “Archiving the Web”, 2007
6 See The Powerhouse Museum’s Electronic Swatchbook or University of Pennsylvania Library’s PennTags.
7 Charles van den Heuvel, “Web Archiving in research and historical global collaboratories. Interfaces and annotations to web archives in the humanities and social sciences seen from a historical perspective” (paper presented at Web_site Histories - Theories, Methods, Analysis, Aarhus, Denmark, October 14, 2008). In N. Brügger ed. Web History, publication foreseen in the book series Digital Formation [ed. Steve Jones] by Peter Lang Publishing Group.
8 Paul Otlet. “Le Livre Dans Les Sciences” (paper presented at the Conférence faite à la maison du Livre, Bruxelles, Musée du Livre, fasc 25-26 1913), 385.
9 Paul Otlet. “Réseau Universel De Documentation: Documentatio Universalis (Rete).” *IID Communicationes* 4(1937): 13-16.
10 W. Boyd Rayward, “Visions of Xanadu: Paul Otlet (1868-1944) and Hypertext,” In *Historical Studies in Information Science*, edited by T. Bellardo Hahn and M. Buckland, 65-80. Medford, NJ: Information Today Inc, 1998, 71.
11 Ibid., 71.
12 Paul Otlet, *Monde, Essai D’universalisme : Connaissance Du Monde, Sentiment Du Monde, Action Organisée Et Plan Du Monde*. Bruxelles: Editiones Mundaneum/D.van Keerberghen &Fils, 1935, 238.
13 Ibid., 238-239.
14 Paul Otlet, *Traité De Documentation : Le Livre Sur Le Livre : Théorie Et Pratique*. Bruxelles: Editiones Mundaneum, Palais Mondial, 1934, 319.
15 Ibid., 319.
16 Alex Wright, *Glut: Mastering Information Through the Ages*. Washington, D.C.: Joseph Henry Press, 2007, 192.
17 Otlet, “Réseau Universel De Documentation: Documentatio Universalis (Rete),”1937.
18 Birger Hjörland and Hannah Albrechtsen, “Toward a New Horizon in Information Science: Domain Analysis.” *Journal of the American Society of Information Science* 46(1995): 400-425.
19 Chuck Zerby, *The Devil's Details: A History of Footnotes*. New York, NY: Touchstone, 2003, 144.
20 Jenny Lynn Bader, “Forget Footnotes. Hyperlink. Old Media, Meet New Media.” *New York Times*, July 16 2000, 1.
21 American Council of Learned Societies, *Our Cultural Commonwealth ,* 2006, 11.
22 Steven M. Schneider, et al., “Web Archiving in E-Research,” in *E-Research: Transformation in Scholarly Practice*, edited by Nick Jankowski, London: Routledge, 2009.
23 Paul Wouters, “Forging New Academic Identities: Key Dilemmas and Challenges in Digital Scholarship.” Paper presented at the University of Washington Simpson Center, Seattle, WA, May 1 2007.
24 Emanuele Quintarelli et al., “Facetag: Integrating Bottom-Up and Top-Down Classification in a Social Tagging System.” Paper presented at the IA Summit, Las Vegas, NV, March 22-26 2007.
25 J. Hunter et al., “HarVana - Harvesting Community Tags to Enrich Collection Metadata.” (paper presented at the Joint Conference on Digital Libraries, Pittsburgh, PA, June 16-20 2008.
26 Robert Anderson, et al., Gentech: Genealogical Data Model Phase 1: A Comprehensive Data Model for Genealogical Research and Analysis, 2000, available from http://xml.coverpages.org/genealogy.html#gentech; accessed April 9, 2009.