

Web Sphere Analysis: An Approach to Studying Online Action

Forthcoming in **Virtual Methods: Issues in Social Science Research on the Internet**, Christine Hine (Ed.), Berg Publishers, Oxford, In Press.

Steven M. Schneider
SUNY Institute of Technology
steve@sunyit.edu

Kirsten A. Foot
University of Washington
kfoot@u.washington.edu

Online action – what people do (or don't do) – alone and together on the World Wide Web and via other Internet applications, is drawing the attention of a wide range of social researchers. The World Wide Web (hereafter referred to as 'the web') can be viewed as an evolving set of structures supporting online action, which manifests and enables the production, inscriptions and experience of cyberculture-- with a myriad of social, political, and cultural dimensions. The hyperlinked, co-produced and evolving characteristics of the web necessitate reconsideration of traditional research methods, and the development of new ones. Each of these characteristics poses particular challenges for researchers. For instance, the hyperlinked and multi-level nature of the web makes the identification and demarcation of units of analysis a critical but difficult task. Seemingly straightforward questions, such as what constitutes a web site, and from what or whose perspective (i.e. robot, browser, or human) that question will be framed, require careful consideration. The co-produced nature of the web, evidenced in the joint production by multiple actors of many features and much content, makes problematic the attribution of agency to producers of specific bits. The often rapid and unpredictable evolution of the web is one of the greatest challenges scholars face as they seek to develop methodological approaches permitting robust examination of web phenomena over time

Scholars from a variety of disciplines are interested in analyzing patterns within and across web materials— some in order to document and make sense of web-based phenomena, others to understand relationships between these patterns and factors exogenous to the web. We suggest online action can be explored, and at least partially explained, through an examination of web objects. These objects, including texts, features, links and sites, can be viewed both as inscriptions of web producers' practices and as potentiating structures for online action on the part of web users. In this approach, web objects and the technologies used to create them are considered as tools that are employed in and that mediate these practices as well as artifacts resulting from them.

In this chapter we describe an approach to research on online action that we call web sphere analysis. This approach centers on the concept of a web sphere as a unit of analysis, and provides

an integrative framework for structural, rhetorical, and sociocultural methods of analysis. We propose three dimensions of web spheres—anticipatability, predictability, and stability-- and explain the implications of various web sphere characteristics for the study of online action within and across them. Finally, we assess the affordances and challenges of web archiving for the purpose of web sphere analysis. We highlight the interrelated set of choices facing researchers as they consider the dimensions of web spheres, select possible methods of analysis for analyzing online action within them, and evaluate techniques employed in identifying and archiving these web spheres.

Web Sphere Analysis

Web sphere analysis is a framework for web studies that enables analysis of communicative actions and relations between web producers and users developmentally over time (Foot and Schneider, 2002, Foot et al., 2003a). We conceptualize a web sphere as not simply a collection of web sites, but as a set of dynamically defined digital resources spanning multiple web sites deemed relevant or related to a central event, concept or theme, and often connected by hyperlinks. The boundaries of a web sphere are delimited by a shared topical orientation and a temporal framework. A significant element in our conceptualization of web sphere is the dynamic nature of the sites to be included. This dynamism comes from two sources. First, the researchers involved in identifying the boundaries of the sphere are likely to continuously find new sites to be included within it. Second, the definition of a web sphere is recursive, in that pages that are referenced by other included sites, as well as pages that reference included sites, may be considered as part of the sphere under evaluation. Thus, as a web sphere is analyzed over time its boundaries may be dynamically shaped by both researchers' identification strategies and changes in the sites themselves.

December's (1996) typology of units of analysis for Internet-related research is useful as a framework for understanding the nature of a web sphere as a unit of analysis. The five types of units of analysis December identifies are: (1) media space, consisting of the set of all servers of a particular type that may provide information in one or more protocols, the corresponding clients that are capable of accessing these servers, and the associated content available for access on these servers; (2) media class, a particular set of content, servers, and clients; (3) media object, a specific unit in a media class with which the user can observe and interact; (4) media instance, a media object at a particular time; and (5) media experience, a particular user's perception of a set of media instances. In correspondence with December's definitions of each kind of unit of analysis, a web sphere could be considered a subset of an Internet media space, constituted by a single media class (web sites), which are comprised of elements or objects such as links, features and texts that are combined in larger media objects of web pages and web sites. What distinguishes our concept of a web sphere from December's definition of a web space is the addition of a shared thematic or event orientation and a temporal framework that is associated with "the set of all Hypertext Transfer Protocol (HTTP) servers, web clients, and content on HTTP servers" that December (1996) identifies.

The web sphere can function as a macro, aggregate unit of analysis, by which historical and/or inter-sphere comparisons can be made. For example, the web sphere of the 2000 elections in the United States can be comparatively analyzed with the electoral web sphere of 2002 and those that develop in later years. Similarly, this web sphere could be contrasted with electoral web spheres in other countries. Other, more micro and/or molar units such as a text, feature, link, site-- or the multi-site web presence of an actor-- can be employed in analyses simultaneously within a web sphere (e.g. Schneider and Foot, 2002, Schneider and Foot, 2003). Defining any of these units operationally can be challenging, particularly when the temporal and malleable aspects of web objects are considered. For example, since any web text or feature can appear stable but actually be modified by its producer and/or rendered differently by technologies such as web browsers employed by users at a particular moment, the point in time and the way in which a web object is observed must be part of the unit's definition for research purposes. Units such as an actor's web presence must also reflect the potential for change over time, by being situated in a particular temporal period. For instance, the web presence of a political party might be appropriately specified by the particular week or month of an election cycle.

Web sphere analysis provides a framework for investigating relations between producers and users of web materials as potentiated and mediated by the structural and feature elements of web sites, hypertexts and the links between them. In its fullest form, the multi-method approach of web sphere analysis consists of the following elements. Web materials related to the object or theme of the sphere are identified, captured in their hyperlinked context, and archived with some periodicity for contemporaneous and retrospective analyses. The identified constituent elements are annotated with human and/or computer-generated "notes" and/or codes of various kinds, which creates a set of metadata. These metadata correspond to the unit(s) and level(s) of analysis anticipated by the researcher(s). Sorting and retrieval of the integrated metadata and URL files is accomplished through several computer-assisted techniques. Interviews, focus groups, experiments and/or surveys are conducted with producers and users of the web sites in the identified sphere, to be triangulated with web media data in the interpretation of the sphere.

Characterizing Web Spheres

Having defined the concept of a web sphere, and given a brief overview of web sphere analysis, we now turn to an explanation and illustration of web sphere dimensions, and an exploration of the challenges associated with discovering/establishing a web sphere and with analyzing online action in the context of a web sphere. As Hine (2000) and Jones (1999) argue, Internet phenomena are in some ways extant for researchers to discover and in other ways constituted by the activity of any user, including a researcher. Thus, the first step in web sphere analysis-- demarcating boundaries and identifying elements-- is both a process of discovery and a process of creation. We identify three dimensions of web spheres that have bearing on how researchers demarcate a web sphere: thematic anticipatability, actor predictability and the stability of constituent web materials. The position of a web sphere of interest on these three dimensions, each of which we view as a continuum, may help researchers develop strategies for identification of elements to be included in the analysis. We should caution that these three dimensions are

neither exhaustive nor mutually exclusive; they represent a starting point for the characterization of this new unit of analysis.

The first dimension measures the degree to which it is possible to *anticipate the emergence* of a web sphere. Part of the anticipatability of a web sphere is dependent on the extent to which it is defined by a specific event; some events, like elections and the Olympics, are highly anticipated, while others, such as accidents, tragedies and scientific discoveries, are less likely to be anticipated. Triggering events (Gamson and Modigliani, 1989), which in general are unanticipated, may also provide the stimulus for a thematic web sphere. For example, a web sphere focused on cloning may have emerged in response to the announcement of Dolly, the first cloned mammal, suggesting that web spheres may emerge in response to specific events rather than be focused on specific events. In the absence of generalized and systematic web archiving (Kahle, 1997, Schneider et al., 2003), anticipatability is often a crucial factor in whether a web sphere is researched. Web spheres emerging quickly after an unanticipated event may be more difficult to study, as a rapid investment of resources (e.g. time, money, topical expertise) may be required. At the same time, a research design can be explicitly tailored to account for analysis of a web sphere “in progress.”

A second dimension of web spheres is concerned with the ability of researchers to *predict the types of actors* who will produce materials encompassed within a web sphere in advance of its emergence. Some web spheres will be produced by a highly predictable set of actor types. Web spheres organized around electoral campaigns, for example, may (depending on the localized political context) include sites produced by parties, candidates, press organizations, advocacy groups, citizens and government agencies. Web spheres emerging around unanticipated natural disasters and accidents are likely to include sites produced by a predictable set of actors: government agencies, relief and charity organizations, press organizations, and citizens, for example. Other web spheres will be produced by a less predictable set of actor types. Following the terrorist attacks of September 11, 2001, we observed significant and unpredicted activity on web sites produced by corporations and businesses, along with more predictable activity on sites produced by religious organizations, educational institutions and government agencies. Actor predictability can greatly affect the thoroughness with which relevant sites can be identified. A predictable set of actors enables researchers to identify a universe of sites to examine for evidence of web activity within a demarcated web sphere. A less predictable set of actors makes this task more difficult, and requires additional searching and identification activities.

Our third dimension is the *level of stability* in the development of sites, links, and other objects within the web sphere. The level of stability has an impact on how frequently the boundaries of the web sphere need to be reconsidered, and how often the sites within the web sphere need to be examined. We have identified three determinants of the level of stability within a web sphere. First, consideration should be given to the frequency of entry and exit of new producers. The extent of new entrants into the web sphere (represented by specific producers, rather than types of producers), and the extent to which producers stop updating, maintaining or serving their sites, is a measure of the stability of the sphere; highly stable spheres have less entry and exit than unstable spheres. Second, stability is a measure of the degree to which sites being analyzed

change or add links to other web sites that ought to be considered within the sphere; frequent changes and substantial additions reduce stability. Third, the frequency and breadth of changes to content and features within the web sites under examination contributes to stability; highly stable web spheres include sites with infrequent and narrowly focused changes to content and features.

Of course, the object of some research projects may be to measure stability and/or to test assumptions about predictability, thus making it problematic to use the phenomenon under consideration as part of the selection process for objects to be studied. Nevertheless, prior research, preliminary examination, and incremental evaluation can contribute both to estimates of these dimensions of the web sphere and to the research design itself.

The position of the web sphere on these dimensions may have bearing on whether the researcher “fixes” the boundaries of the sphere at the beginning of a study, or engages in a dynamic bounding process. Bounding refers to the process of identifying constituent elements (sites or pages) within the web sphere and specifying a temporal frame for analysis. Identifying constituent elements of the web sphere to be examined may include both establishing the universe of sites or pages about which generalizations can be offered, and specifying a method of sampling sites or pages to be analyzed. Constituent elements can be identified prior to analysis – following long-established practices in survey research (Hyman, 1955) and content analysis (Berelson, 1952). Alternatively, constituent elements can be identified as part of the analysis process, building on well-established techniques used in participant observation research (Whyte, 1943), and more recently labeled as a snowball research strategy (Berg, 1988, Atkinson and Flint, 2001). Identifying constituent elements prior to analysis – in effect, fixing the boundaries of the web sphere under study – offers several advantages to the researcher. A clearly defined universe of sites within a web sphere makes representative sampling of sites possible for both archiving and structured observations. Fixed boundaries may also increase the possibility of replicating findings by subsequent analysts. Finally, fixing the boundaries of the web sphere may enhance options for collaboration in archiving or analyzing the web sphere, particularly with entities such as libraries (Schneider et al., 2003).

On the other hand, dynamic bounding allows the researcher to be responsive to unanticipated developments and emergent trends in the web sphere. Prior research in both the September 11th web sphere and electoral web spheres suggests that, even within anticipated, predictable and generally stable web spheres, unanticipated events precipitate the production or alteration of intertextual and interlinked web objects, sometimes in a matter of hours or over the course of a few days. We employ the concept of a web storm as a unit of analysis that reflects inter-actor and inter-site activity over a relatively brief period of time. For instance, a political scandal is likely to result in a web storm wherein actors such as news organizations, advocacy groups, and individual citizens, (e.g. producers of politically-oriented weblogs), post texts, graphics and links regarding the scandal intensively for several days or weeks. Some web storms develop into web spheres that are durable on the web over a longer period, often through the migration of individual texts and pages pertaining to an event onto sites newly produced and dedicated to the event. For example, a web storm emerged quickly in the wake of the release of the Starr report

detailing U.S. President Bill Clinton's affair with Monica Lewinsky and raising questions of perjury. The web storm of commentaries developed into a web sphere as sites were produced to advocate impeachment of the president, (e.g. impeachclinton.org¹), or oppose it, (e.g. <http://moveon.org>²).

Unless a researcher is engaged in dynamic bounding, he or she is likely to miss the opportunity to analyze these web storms, which may be significant bursts of online action. Dynamic bounding as a scholarly practice is also more consistent with how the web functions from a user perspective. It is critical, though, that dynamic bounding be implemented systematically to ensure representativeness and replicability.

Completely fixed and fully dynamic bounding represent two ends of a spectrum for identifying constituent elements within web spheres—most researchers will choose some blend of the two. Researchers need to decide, preferably before beginning a study, how frequently the web sphere boundaries will be re-defined, and under what circumstances, and how often and according to what criteria and techniques the web will be searched for pages, sites or links comprising the web sphere. A further factor to consider is the correspondence between research goals and the bounding strategy. For instance, if the goal is to analyze the development of a web sphere, a fairly dynamic bounding strategy is needed. Finally, researchers should assume that increased dynamism in web sphere demarcation will increase resource needs—time, effort, and storage – especially if systematic archiving is involved.

The process of discovering/establishing the web sphere under study also includes defining the steps or procedures to be taken to identify the specific elements to be examined. Depending on the characteristics of the web sphere, this process can involve a number of strategies. If the producer types are highly predictable, and the web sphere itself highly anticipated, existing and maintained directories of sites may be available. For example, if a researcher was interested in the web sphere related to a single season of a professional sports league, a pre-existing directory of sites representing each of the participating teams could likely serve as a starting point for identification of relevant sites. Encompassing the pre-existing directory into the research design (i.e. by specifying sites to be examined as those identified by the directory on a specific date) both serves as a robust specification of identification strategy and provides a universe from which samples of sites to be examined can be reliably drawn.

On the other hand, some web spheres of interest, especially those that are not anticipated or predictable, may require a more search-oriented strategy to identify constituent elements. Using topical key words systematically in search engines may be fruitful in identifying constituent elements of a web sphere. However, this strategy, which relies on the presence of “relevant” content within potential sites, may have significant drawbacks. The absence of relevant content on some web sites may reflect action on the part of a site producer that is just as strategic as the presence of relevant web content would be. Furthermore, if relevant material were to appear on a particular site at a later date, establishing its absence at the beginning of the study period would be critical for developmental analyses.

One alternative to content relevance as the primary criteria for inclusion in the web sphere is the identification of relevant actor types, and the inclusion of web sites produced by identified actors. Relevant actor types can be determined in several ways, beginning with whatever extant literature informs a particular study, as well as through methods of social network analysis. Once relevant actor types are established, the web sites of particular actors within each type can be selected through various indices and sampling techniques. Web sites produced by relevant actors may be significant in structuring online action-- or the lack thereof-- within the web sphere even if those actors have not (yet) produced web materials relevant to the theme of the sphere at the beginning of the study.

Another strategy for identifying constituent elements in a web sphere is to analyze patterns of inlinking to and outlinking from a core set of URLs. For instance, the web sphere of a sports team could be defined by tracing the inlinks to and outlinks from the URLs of the home pages of each player and the team itself. The origin pages of inlinks and the destination pages of outlinks can be analyzed in the context of their base sites to identify their producers, then to specify the producer type. Further analysis of these producers' web presence may be helpful to ascertain their position or stance in the web sphere.

Approaches to Studying Online Action in a Web Sphere

The processes of demarcating/establishing web spheres, and of identifying constituent elements in a web sphere are foundational to tracking developmental trajectories of online action. In this section, we suggest a range of methods that can be adapted for studying online action within the integrative framework of web sphere analysis by identifying three sets of approaches that have been used in web-related research over the last decade (Schneider and Foot, In Press). Although these approaches are not necessarily mutually exclusive, distinguishing between them enables examination of the affordances and challenges of each for studies of online action within a web sphere.

The first set of approaches employ *discursive or rhetorical* analyses of web pages or sites. Typically, this set of approaches is more concerned with the content of a web site than its structuring elements. Studies employing these approaches focus on the texts and images contained on web pages, and/or on web pages/sites as texts in a Foucauldian sense (e.g. Baym, 1999, Benoit and Benoit, 2000, Sillaman, 2000, Warnick, 1998). In an online action perspective, web texts are situated as inscriptions of communicative practices on the part of site sponsors and/or users, and methods of discursive or rhetorical analysis can help to illuminate social action. For example, discursive and rhetorical analyses were employed in the context of analysis of the post-September 11 web sphere in studies of personal expression in general (Siegl and Foot, In Press), and memorializing in particular (Foot et al., Under review) as forms of online action. As further illustration of the potential role of this set of approaches in web sphere analysis, rhetorical analysis was combined with quantitative content analysis of issue stance texts on campaign sites from 200 races to study campaign practices of position taking online (Foot et al., 2003b). Discursive or rhetorical methods can also shed light on relations between web producers within a

web sphere, particularly when hypertext intertextuality is included through analyses of cross-site linking, (e.g. Mitra, 1999, Warnick, 2001).

The second set of approaches are *structural/feature* analyses. Studies in this genre that use individual web sites as the unit of analysis focus on the structure of the site, such as the number of pages, hierarchical ordering of pages, or on the features found on the pages within the site, for instance, the presence of a search engine, privacy policy, or multiple navigation options (D'Alessio, 1997, D'Alessio, 2000, McMillan, 1999, Hansen, 2000, Benoit and Benoit, 2000). Other types of structural analysis employ computer-assisted macro-level network analysis methods for mapping linking patterns (Jackson, 1997, Park, 2003, Park and Thelwall, 2003, Rogers and Marres, 2000, Rogers and Marres, 2002, Thelwall, 2001, and Park and Thelwall, this volume). Studies of this type enable understanding of network structures on the web, but inferring the meaning or “substance” of those network structures can be difficult to infer from large-scale mapping studies. We view features and other structural elements of sites as manifestations of (co)production activities that constitute “online structure” (Schneider and Foot, 2002). Online structure both inscribes particular forms of communicative, social, and/or political action on the part of web producers, and enables or constrains particular forms on online and offline action on the part of web users. For example, a campaign’s decision to create a database-driven web site feature for sending electronic letters to local newspaper editors enables online political mobilization. Methods for systematically analyzing features and other types of structure within a web sphere are being developed and employed in studies of online action both alone (e.g. Foot et al., 2003a) and in combination with other methods that elicit user perspectives such as focus groups (e.g. Stromer-Galley and Foot, 2002) and surveys of web users (e.g. Stromer-Galley et al., 2001, Schneider and Foot, 2003).

A third set of approaches to web analysis has emerged for analyzing multi-actor, cross-site action on the web, based on adaptations of sociocultural methods of inquiry. Lindlof and Shatzer (1998) point in this direction in their article calling for new strategies of media ethnography in “virtual space.” Hine (2000) presents a good example of sociocultural analysis of cross-site action on the web. Similarly, Howard’s (2002) conceptualization of network ethnography reflects methodological sensitivity to processes of web production as a form of strategic online action. By appropriating the term “sociocultural” to describe this set of approaches we seek to highlight the attention paid in this genre of web studies to the hyperlinked context(s) and situatedness of web sites-- and to the aims, strategies and identity-construction processes of web site producers (see several examples in Beaulieu and Park, 2003, and Beaulieu, this volume). Field research methods of participant observation and interviews, in combination with textual, structural and link analyses comprise this set of approaches. For studies of online action within a web sphere, sociocultural approaches are particularly helpful in analyzing complex, evolving processes such as collaborative mobilization and coproduction-- of features, pages, and whole sites, and across sites through links (Foot and Schneider, 2002, Forte, 2003, and Forte, this volume).

In summary, the rubric for selecting and adapting social research methods for the study of online action is comprised of theoretical perspectives that situate web productions as socio-technical formations in combination with research design choices regarding levels and units of analysis.

Within the framework of web sphere analysis, a wide range of methods can be adapted and used fruitfully in various combinations, in correspondence with particular kinds of research questions regarding online action. All of the methods described above will result in the generation of some kind of annotations or metadata about web materials, whether qualitative or quantitative. As we discuss in the next section, researchers may find it useful to create an archive or database of web materials in conjunction with their database of metadata.

Archiving Materials for Web Sphere Analyses

Methodological choices implicated in the identification of a web sphere and the selection of research methods for analyzing online action, have been discussed above. The third methodological issue entailed in analyzing online action within the framework of web sphere analysis concerns collecting or archiving “born digital” web materials as data. Research design decisions about whether and how to archive for studies of online action may be influenced by the dimensions of the web sphere under consideration, and by the approach(es) selected for analyzing online action. For instance, in-depth rhetorical analysis of a relatively small number of web texts produced within a stable and predictable web sphere during a relatively brief period would not require the same level of archiving as a structural analysis of the development of a large number of web sites in an unstable and unpredictable web sphere over the course of several months. In this section, we focus on technical choices that may be involved in archiving for research projects within the framework of web sphere analysis in view of the nature of web materials, and on the specific challenges researchers face in collecting these materials as data.

Web materials are uniquely ephemeral and persistent. The ephemerality of web content comes from both its transience and its construction. The availability of web content, once produced, is often solely under the control of the producer-- a characteristic of born digital materials that holds significant implications for scholarly citation as well as data collection (Weiss, 2003). From the perspective of web users, including social researchers, specialized tools and techniques are required to ensure that content can be viewed again at a later time. Additionally, web content, like theater and other “performance media” (Hecht et al., 1993, Stowkowski, 2002), is ephemeral in its construction: once presented, it needs to be reconstructed or re-presented in order for others to experience it. Although web pages are routinely reconstructed by computers without human intervention, (i.e. when a request is forwarded to a web server), it nevertheless requires some action by the producer, or the producer’s server, in order for the content to be reproduced or reconstructed. Put another way, the experience of the web, as well as the bits used to produce the content, must be intentionally preserved in order for it to be reproduced (Arms et al., 2001). This is in contrast to older media – including printed materials, film and sound recordings, for example – that can be archived in the form in which they are presented; without additional steps to re-create the experience of the original.

At the same time, the web has a sense of persistence and even permanence that clearly distinguishes it from performance media. Unlike theater, or live television or radio, web content must exist in a permanent form in order to be transmitted. Even “dynamically generated” web

pages generally rely on data previously encoded in databases.³ The web shares this characteristic with other forms of media such as film, print, and sound recordings. The permanence of the web, however, is somewhat fleeting. Unlike any other permanent media, a web site may regularly and procedurally destroy its predecessor each time it is updated by its producer. That is, in the absence of specific arrangements to the contrary, each previous edition of a web site may be erased as a new version is produced. By analogy, it would be as if each day's newspaper was printed on the same piece of paper, obliterating yesterday's news to produce today's.

Researchers interested in studying online action may need to counter this ephemerality by taking pro-active steps in order to facilitate a recreation of web experience (Arms et al., 2001) for future analyses. The permanence of the web makes this eminently possible. Although saving web sites is not as easy as, say, saving editions of a magazine, archiving techniques are evolving in such a way to facilitate scholarly research of web sites. In distinction to other ephemeral media, the web can be preserved in nearly the same form as it was originally "performed" (Kahle, 1997, Lyman and Kahle, 1998, Lyman, 2002) and analyzed at a later time.

Decisions about archiving strategies should be made with reference to research design, research operations, and ethical considerations. The research design should include a specification of periodicity, breadth, and depth. Periodicity refers to the frequency with which the researcher desires to capture or archive the web materials under examination. Depending on the stability of the materials, the researcher's objectives with respect to conducting developmental analyses measuring change over time, and the time frame anticipated for the research project, researchers may wish to capture materials daily, weekly, monthly, quarterly or even annually. Breadth in this context refers to the inclusion (or exclusion) of pages that are hyperlinked to the sites or pages under study. Pages with links to the sites under examination ("inlinks"), as well as pages that are linked by the sites under examination ("outlinks"), may provide critical context for understanding or evaluating the online action within the web sphere being analyzed. One approach is to include pages one link away from those sites identified as being in the core of the web sphere. Inclusion of inlinked pages will, by definition, require additional searching for identification purposes. Inclusion of outlinked pages may present significant challenges for some archiving tools. Depth references the number of levels from a "base" or starting web page, within the same domain or web site, that the researcher wishes to include within the web sphere. Many researchers may prefer to capture the whole web site without regard to depth. Periodicity may also interact with breadth and depth. For example, archiving strategies can be designed to capture "deep" materials with a less frequent periodicity -- perhaps capturing front pages of sites daily, and the second and subsequent levels of sites weekly or monthly.

Research operations concern the techniques that will be employed to conduct the analysis. First, the sheer quantity of materials to be archived will likely affect the techniques developed -- capturing and maintaining an archive of several hundred gigabytes presents challenges substantially different than an effort involving several dozen gigabytes. Second, the number of individual researchers involved in the project may dictate specific choices. A collaborative project involving two or more researchers may require the use of a centralized server to host the archived materials, and the implementation of a system to allow web access to the archive.

Alternatively, research projects may function quite well with single copies of archived materials stored on removable media like CDs or DVDs. Systems to track references to specific instances of archived objects (sites, pages, etc.) may also be dependent on the method developed to provide access to archived materials

Finally, archiving decisions should also be made with consideration of the intentions concerning the long-term use and availability of the captured materials. Scholars ought to consider the ethical issues associated with archiving (Schneider et al., 2003), such as current copyright laws, the phenomenon of artifacts being produced in the archiving process that were never actually available on the web, and the potential for digital repositories to be exploited in harmful ways by current and future users. Researchers may find it necessary to consult institutional policies concerning these issues, especially those involving copyright, fair use and consent. In addition, researchers may wish to consider the value of their archived materials to other scholars and to the wider community. Social science data archives and libraries represent two potential repositories for these materials. Choices made when capturing materials – for example, including HTTP header information -- may increase the likelihood that archives match standards established by these types of institutions.

Conclusion

Web sphere analysis holds potential as an approach for studying online action. This approach is centered on the web sphere as a clearly-specified, multi-dimensional unit of analysis. It provides an integrative framework encompassing multiple research methods. We have made explicit the set of challenges involved in identifying and collecting data responsive to the selected research method(s) and characteristics of the web sphere under examination. The framing of the web sphere, selection of research method, and process of identifying and collecting data represent a set of interrelated choices to be made by researchers during the research design process.

Researchers interested in analyzing a web sphere should explicitly define its boundaries and the procedures to be employed in dynamically shaping those boundaries during the course of the study. Reflection on the three dimensions of a web sphere – the anticipatability of its emergence, the predictability of the types of actors producing materials, and the level of stability – may contribute to decisions made as researchers approach a web sphere analysis. In particular, the dimensions of the web sphere under study may shape the data collection strategy.

Three approaches to studying online action illustrate the range of analysis methods that can be employed. Discursive or rhetorical analyses are more concerned with the content of web sites than its structuring elements. Structural/feature analyses focus on the structure of sites, and may employ macro-level network analysis methods for mapping relations between sites. Sociocultural approaches focus on the hyperlinked context(s) and coproductive aspects of sites, features and texts, and of cross-site action through links. The type of web sphere that is under investigation has implications for the selection of one or more research methods, and may shape both the web sphere identification and data collection processes.

Data collection processes, which include both archiving web materials and the creation of meta-data characterizing web materials, are highly dependent on a number of factors. The characteristics of the web sphere under examination and the research method selected will likely affect data collection. In addition, these processes will be affected by the availability of resources such as archiving expertise, computing facilities, time frame, and research support. Web archiving is particularly useful for developmental analyses that are time sensitive (e.g. Foot et al., 2003a), and in analyses of any kind in unstable and unpredictable web spheres. Regardless of the analytical methods employed, or the dimensions of the web sphere under investigation, web archiving enables more rigorous and verifiable research. Finally, consideration of ethical, policy and legal factors may affect archiving and data collection decisions.

The study of online action will continue to draw the attention of social researchers. Emerging practices of web production and use evoke questions about the traditional distinctions between producers and users. Casting this set of questions within the approach of web sphere analysis offers researchers a framework to productively and robustly conceptualize analyses of the hyperlinked, co-produced set of structures that manifest and enable the production, inscriptions and experience of cyberculture.

References

- Arms, W., Adkins, R., Ammen, C. and Hayes, A. (2001) 'Collecting and Preserving the Web: The MINERVA Prototype', in *RLG DigiNews*, 5 (2).
<http://www.rlg.org/preserv/diginews/diginews5-2.html> (Accessed 25 February 2004)
- Atkinson, R. and Flint, J. (2001) 'Accessing Hidden and Hard-to-Reach Populations: Snowball Research Strategies', *Social Research Update*, Summer 2001(33).
<http://www.soc.surrey.ac.uk/sru/SRU33.html> (Accessed 25 February 2004)
- Baym, N. (1999) *Tune In, Log On: Soaps, Fandom and On-Line Community*, Thousand Oaks, CA: Sage.
- Beaulieu, A. and Park, H. W. (Eds.) (2003) *The Form and the Feel: Combining Approaches for the Study of Networks on the Internet*, 8(4). <http://www.ascusc.org/jcmc/vol8/issue4/> (Accessed 25 February 2004)
- Benoit, W. J. and Benoit, P. J. (2000) 'The Virtual Campaign: Presidential Primary Websites in Campaign 2000', *American Communication Journal*, 3(3).
<http://acjournal.org/holdings/vol3/Iss3/curtain.html#4> (Accessed 25 February 2004)
- Berelson, B. (1952) *Content Analysis in Communication Research*, Glencoe, IL: Free Press.
- Berg, S. (1988) 'Snowball Sampling', In *Encyclopaedia of Statistical Sciences*, Vol. 8 Kotz, S. and Johnson, N. L. (ed.,^eds.), New York: John Wiley.

D'Alessio, D. (1997) 'Use of the World Wide Web in the 1996 U.S. Election', *Electoral Studies*, 16(4): 489-500.

D'Alessio, D. (2000) 'Adoption of the World Wide Web by American Political Candidates, 1996-1998', *Journal of Broadcasting and Electronic Media*, 44(4): 556-568.

December, J. (1996) 'Units of Analysis for Internet Communication', *Journal of Computer Mediated Communication*, 1(4). <http://www.ascusc.org/jcmc/vol1/issue4/december.html>
(Accessed 25 February 2004)

Foot, K. A. and Schneider, S. M. (2002) 'Online Action in Campaign 2000: An Exploratory Analysis of the U.S. Political Web Sphere', *Journal of Broadcasting & Electronic Media*, 46(2): 222-244.

Foot, K. A., Schneider, S. M., Dougherty, M., Xenos, M. and Larsen, E. (2003a) 'Analyzing Linking Practices: Candidate Sites in the 2002 U.S. Electoral Web Sphere', *Journal of Computer-Mediated Communication*, 8(4). <http://www.ascusc.org/jcmc/vol8/issue4/foot.html>
(Accessed 25 February 2004)

Foot, K. A., Warnick, B. and Schneider, S. M. (Under review) 'Web-Based Memorializing After September 11: Toward a Conceptual Framework', *New Media & Society*.

Foot, K. A., Xenos, M. and Schneider, S. M. (2003b) 'Online Campaigning in the 2002 U.S. Elections: Analyzing House, Senate and Gubernatorial Campaign Web Sites', paper presented at American Political Science Association, Philadelphia, August 28-31.
<http://politicalweb.info/preElection.html#> (Accessed 25 February 2004)

Forte, M. C. (2003) 'Co-Construction and Field Creation: Website Development as both an Instrument and Relationship in Action Research', In *Virtual Research Ethics: Issues and Controversies*, Buchanan, E. (ed.), pp. 222-248. Hershey, PA: Idea Publishing Group.

Gamson, W. A. and Modigliani, A. (1989) 'Media Discourse and Public Opinion on Nuclear Power: A Constructionist Approach.' *American Journal of Sociology*, 95(1): 1-37.

Hansen, G. (2000) 'Internet Presidential Campaigning: The Influences of Candidate Internet Sites on the 2000 Elections', paper presented at National Communication Association, Seattle, WA, November.

Hecht, M. L., Corman, S. R. and Miller-Rassulo, M. (1993) 'An Evaluation of the Drug Resistance Project: A Comparison of Film versus Live Performance Media', *Health Communication*, 5(2): 75-88.

Hine, C. (2000) *Virtual Ethnography*, Thousand Oaks, CA: Sage.

Howard, P. (2002) 'Network Ethnography and Hypermedia Organization: New Organizations, New Media, New Myths', *New Media & Society*, 4(4): 550-574.

Hyman, H. H. (1955) *Survey Design and Analysis: Principles, Cases, and Procedures*, Glencoe, IL: Free Press.

Jackson, M. (1997) 'Assessing the Structure of the Communication on the World Wide Web', *Journal of Computer-Mediated Communication*, 3(1).

<http://www.ascusc.org/jcmc/vol3/issue1/jackson.html> (Accessed 25 February 2004)

Jones, S. (Ed.) (1999) *Doing Internet Research: Critical Issues and Methods for Examining the Net*, Thousand Oaks: Sage.

Kahle, B. (1997) 'Preserving the Internet', *Scientific American*, 276(3): 82-83.

Lindlof, T. R. and Shatzer, M. J. (1998) 'Media Ethnography in Virtual Space: Strategies, Limits, and Possibilities', *Journal of Broadcasting and Electronic Media*, 42(2): 170-189.

Lyman, P. (2002) 'Archiving the World Wide Web', in *Building a National Strategy for Digital Preservation*, Council on Library and Information Resources report.

<http://www.clir.org/pubs/reports/pub106/web.html> (Accessed 25 February 2004)

Lyman, P. and Kahle, B. (1998) 'Archiving Digital Cultural Artifacts: Organizing an Agenda for Action', *D-Lib Magazine*. <http://www.dlib.org/dlib/july98/07lyman.html> (Accessed 25 February 2004)

McMillan, S. J. (1999) 'Health Communication and the Internet: Relationships between Interactive Characteristics of the Medium and Site Creators, Content, and Purpose', *Health Communication*, 11(4): 375-390.

Mitra, A. (1999) 'Characteristics of the WWW Text: Tracing Discursive Strategies', *Journal of Computer-Mediated Communication*, 5(1).

Park, H. W. (2003) 'Hyperlink Network Analysis: A New Method for the Study of Social Structure on the Web', *Connections*, 25(1): 49-61.

Park, H. W. and Thelwall, M. (2003) 'Hyperlink Analyses of the World Wide Web: A Review', *Journal of Computer-Mediated Communication*, 8(4).

<http://www.ascusc.org/jcmc/vol8/issue4/park.html> (Accessed 25 February 2004)

Rogers, R. and Marres, N. (2000) 'Landscaping Climate Change: A Mapping Technique for Understanding Science & Technology Debates on the World Wide Web', *Public Understanding of Science*, 9(2): 141-163.

Rogers, R. and Marres, N. (2002) 'French Scandals on the Web, and on the Streets: A Small Experiment in Stretching the Limits of Reported Reality', *Asian Journal of Social Science*, 30(2): 339-353.

Schneider, S. M., Foot, K., Kimpton, M. and Jones, G. (2003) 'Building Thematic Web Collections: Challenges and Experiences from the September 11 Web Archive and the Election 2002 Web Archive', paper presented at European Conference on Digital Libraries Workshop on Web Archives, Trondheim, Norway.

<http://bibnum.bnf.fr/ecdl/2003/proceedings.php?f=schneider> (Accessed 25 February 2004)

Schneider, S. M. and Foot, K. A. (2002) 'Online Structure for Political Action: Exploring Presidential Web Sites from the 2000 American Election', *Javnost (The Public)*, 9(2): 43-60.

Schneider, S. M. and Foot, K. A. (2003) 'Crisis Communication & New Media: The Web After September 11', In *Society Online: The Internet in Context*, Howard, P. N. and Jones, S. (eds.), pp. 137-154. London: Sage.

Schneider, S. M. and Foot, K. A. (In Press) 'The Web as an Object of Study', *New Media & Society*.

Siegl, E. and Foot, K. A. (In Press) 'Expression in the Post-September 11th Web Sphere', *Electronic Journal of Communication*.

Sillaman, L. (2000) 'The Digital Campaign Trail: Candidate Images on Campaign Websites', unpublished master's thesis, Annenberg School of Communication, Philadelphia, PA: University of Pennsylvania.

Stowkowski, P. A. (2002) 'Languages of Place and Discourses of Power: Constructing New Senses of Place', *Journal of Leisure Research*, 34(4): 368-382.

Stromer-Galley, J. and Foot, K. A. (2002) 'Citizen Perceptions of Online Interactivity and Implications for Political Campaign Communication', *Journal of Computer-Mediated Communication*, 8(1). <http://www.ascusc.org/jcmc/vol8/issue1/stromerandfoot.html> (Accessed 25 February 2004)

Stromer-Galley, J., Foot, K. A., Schneider, S. M. and Larsen, E. (2001) 'How Citizens Used the Internet in Election 2000', In *Elections in the Age of the Internet: Lessons from the United States*, Coleman, S. (ed.), pp. 26-35. London: Hansard Society.

Thelwall, M. (2001) 'Extracting Macroscopic Information from Web Links', *Journal of the American Society for Information Science and Technology*, 52(13): 1157-1168.

Warnick, B. (1998) 'Appearance or Reality? Political Parody on the Web in Campaign '96', *Critical Studies on Mass Communication*, 15(3): 306-324.

Warnick, B. (2001) *Critical Literacy in a Digital Era: Technology, Rhetoric, and the Public Sphere*, Mahwah, NJ: Lawrence Erlbaum Associates.

Weiss, R. (2003) 'On the Web, Research Work Proves Ephemeral', *Washington Post*, Washington, DC, A08.

Whyte, W. F. (1943) *Street Corner Society: The Social Structure of an Italian Slum*, Chicago, IL: The University of Chicago Press.

Footnotes

¹ For an archival impression of this site captured shortly after the impeachment proceedings see <http://web.archive.org/web/20000104071229/www.impeachclinton.com/main/webmastersnote.htm>.

² For an archival impression of this site captured during the impeachment proceedings, see <http://web.archive.org/web/19981202172227/http://www.moveon.org/>.

³ Application-based Web features, such as java applets, are not constrained in this way, and may include the presentation or generation of data not previously constructed by the producer.