

3-10-1995

Hypertext and the Docuverse: A Research Memo

Roger A. Lohmann

West Virginia University, roger.lohmann@mail.wvu.edu

Follow this and additional works at: https://researchrepository.wvu.edu/faculty_publications



Part of the [Nonprofit Administration and Management Commons](#), [Public Administration Commons](#), [Scholarly Communication Commons](#), [Scholarly Publishing Commons](#), and the [Social Work Commons](#)

Digital Commons Citation

Lohmann, Roger A., "Hypertext and the Docuverse: A Research Memo" (1995). *Faculty & Staff Scholarship*. 2583.

https://researchrepository.wvu.edu/faculty_publications/2583

This Article is brought to you for free and open access by The Research Repository @ WVU. It has been accepted for inclusion in Faculty & Staff Scholarship by an authorized administrator of The Research Repository @ WVU. For more information, please contact ian.harmon@mail.wvu.edu.

Hypertexts and the Docuverse: A Research Memo

Roger A. Lohmann
West Virginia University

The term docuverse was first developed by Apple Computer in-house guru Allen Kay in the late 1960's. The underlying idea can be traced back a couple of decades earlier, to the visionary Vannevar Bush and the Memex (Bush, 1945). According to Kay, a docuverse is a set of related documents together with the linkages between them. In this paper, a docuverse is conceived as a collection of documents together with the links, ties and bonds that can bring them together into an integrated logical and conceptual whole. Kay who also coined the term hypertext, which refers to an electronic document with existing and anticipated hypermedia technology. have already made possible the construction of an electronic web of all the related documents of social science (and any other academic discipline, professional field or interdiscipline).

The potentials of multiple paths in hypertext documents opens a number of very intriguing possibilities. For example, creation of a universal electronic library (or docuverse) of texts could be separately organized from writer's (or producer's) and reader's (or audience's) viewpoints upon segments of text organized into electronic files and connected by embedded links. While editors, librarians and computer programmers will each have slightly different approaches to this organization, a most important consideration for writers is finding a consistent basis for what constitutes an electronic hypertext document (the more sophisticated equivalent, perhaps, of an ASCII file) in such a docuverse.

Web pages (beginning with the celebrated "home pages") are the basic structural tool of the world wide web, and by implication of the docuverse proposed here. The question inevitably arises of what such a single web page should consist, and whether there is any measure of organization or order other than simple author-preference to guide the design of such pages. The most basic definition proposed here is that "web pages" are synonymous with reasonable-sized text documents: neither too long nor too short, but just right. Ah, but what is just the right length of a text? Many different answers are possible. One traditional answer arising often in the context of social science journal articles has been "about 20 pages", and writers have learned to write accordingly. Thus, we can define an *article* as a semantically-related text consisting of sentences, paragraphs and sections organized as an introduction, one or more body sections, a conclusion, bibliography and perhaps notes. This is likely to be as true in an electronic environment as it has been in the social science literature reaching back at least to the 1870's. (See, for example, the early publications of the National Conference of Charities and Corrections.)

Another possible unit of organization is the section: Regardless of word or page length, articles may consist of a series of linked and cumulative sections beginning with "a beginning" and proceeding in a coherent fashion to a conclusion or ending. As in ordinary writing, sections, in turn, consist of paragraphs, which consist of sentences which consist of words, which consist of letters. The intriguing question which hypertext authoring is currently attempting to deal with is whether the limits of the web page correspond most closely to the article, the section, or the paragraph? (The sentence, word and letter can probably be safely discounted for the moment.)

Is there any level of conceptual organization which makes sense between the level of a universe of individually linked articles and the docuverse as a whole? The collections of a few electronic journals organized along more or less traditional lines notwithstanding, the majority answer in today's internet would appear to be no. To the present, most "electronic journals" are, in fact, merely folders or archives of separate articles saved in portable document format (PDF). There are not yet any direct machine analogs of the physical unities of the journal issue or volume. Such continuous, sequential collections of articles organized by date of publication correspond closely to our conception of the scientific or professional *journal*, divided by years into *volumes* and some lesser division of time (e.g., quarters) into *issues*.

Likewise, articles, as defined, are usually termed *chapters* when they are presented within the larger logical and semantic structures which we call *books*. Books have a number of characteristics, including covers, dust jackets, front-matter, and back-matter like *indexes* and the like which are of great concern to publishers and librarians, but for author and reader. This point is strengthened if we regard *preface*, *introduction*, *table of contents* and other front matter as specialized subordinate texts, along with *abstracts* and end matter like *bibliographies*, rather than as integral parts of the main texts to which they relate. (In an electronic environment, indexing is likely to revert to an operation (e.g., a Google search on the internet) rather than a text.

Thus, the structure of a docuverse can, to some degree, model the traditional practices of scholarly publication and allow for *virtual volumes* like journal issues and books and even for collections of *virtual multi-volumes sets* like the various encyclopedias of social science, economics, sociology, social work, corrections, etc. The external reference statements of HTML allow for extending the appearance of the traditional hierarchical unities of such collections (articles within volumes; volumes within sets; sets within "supersets"; volumes, sets and supersets organized into libraries, libraries into a docuverse; etc.).

What, then are the basic articles (texts, documents or files) out of which a social science docuverse is to be constituted from authors' and readers' perspectives? These appear to fall into two distinct series, termed simply *main* articles and *subordinate* articles. Main articles are those which carry

forward central semantic strains (arguments, presentation of evidence, etc.) and would include original articles of numerous types and peer reviews, both anonymous and signed. Subordinate articles could include such matter as abstracts, notes (whether footnotes or endnotes), citations to other works, bibliographic or reference lists, other works by the author (perhaps, the publications section of an author's vita), biographical sketches of authors, and multi-media illustrative materials.

Given the location of most scholarly research and writing in colleges and universities, it is important to note also a special class of main documents associated not with the traditional activities of publication, but rather with the traditional activities of teaching. Thus, the *syllabus*, with its traditional *topical outline* and *reading list* and other sections could, in an electronic docuverse, be readily linked with other elements of the docuverse as outlined above.

Treating subordinate texts, like those detailed in Appendix A, as separate documents in an electronic docuverse seems desirable because of their "timely" (that is, changing) manner. Including author's publications lists and biographical material within texts is likely to prove only confusing, since such lists change so frequently. By contrast, including constant *links* to such materials would mean that the material could be constantly updated without having to go back and update any links, so long as the URL of the document(s) being revised remained constant.

This recommendation, however, also points to one of the major weaknesses of electronic publishing of the sort suggested here: How to assure the integrity of the system. What does "constant" mean in an electronic context? How, for example, are unethical authors to be prevented from simply downloading the works of others, making minor changes and submitting it in another context under their own names? For that matter, what is to prevent an unscrupulous author, for example, from simply substituting another (and inferior) article under the same URL of an (superior) article already peer-reviewed and accepted for publication?

Links, Ties and Bonds

It may also be desirable in a docuverse of this sort to create one or more special classes of abbreviated URL's to mark links between a particular main article and its subordinate articles, and to set them apart from links between main articles. From the standpoint of HTML (or other subsequent document markup languages), nothing would be different; this is largely a matter of writing conventions and display rather than separate anchors.

Thus, the *ties* between a particular sentence and a particular footnote might be linked with sequential numbers or letters, as in conventional writing practice, and display the footnote text when selected. HTML named anchors make such links easy to construct. Likewise, ties between an in-text

citation (e.g., Author, Date) and the corresponding bibliographic citation, could use the same type of anchors whether the reference list is at the end of the text or in a subordinate article.

HTML makes it possible to write additional, non-exclusive ties which would be logical and useful in the traditional scholarly context. A second class of possible ties are "TOC Anchors", which link lines in a table of contents to specific subheadings or sections within an article. Many on-line authors have already begun to attach what are functionally tables of contents to their on-line articles because of the inherent limits of the screen. (See, for example, the Beginners Guide to HTML at <https://www.htmldog.com/guides/html/beginner> .

It may be useful, or at the very least, interesting, in an electronic context, to restore the more detailed tables of contents found in 18th and 19th century books, in which not only chapter headings, but individual section headings/topics are listed. In any case, the practice of linking such a TOC "header" to appropriate sections within an article would be a useful and desirable practice to enable moving about in a text. Taken together, these ties are all within the logical limits of a main article as such. A second class of links, referred to here as *bonds* , would link a main article and an appropriate set of subordinate articles.

What might be called *context anchors* could bond a bibliographic citation to (perhaps enclosing the entire article title so that it is highlighted) the referent text. This could apply to either direct and indirect quotations and allow a skeptical or critical reader to examine more closely the exact context from which a finding, interpretation or quotation was lifted. The question arises here of whether, in the long run, it might not be preferable to embed such context anchors directly in the text, include full citations and URL's at the beginning of each article, and eliminate the need for the traditional reference list as such.

Five particular subordinate articles will be considered here: abstracts, bibliographies, notes, peer-reviews and commentaries. At present, the first three of these subordinate articles, when they are published at all, tend to be included as sections within articles or books, obscuring to some degree their independence. The abstract is, to a real degree, an independent text with a distinct purpose. Whether written by an author or independent abstractor, its function is to summarize, or focus upon the main points of the article being abstracted. There is a great deal of necessary redundancy in the present paper-and-print system, where an abstract may be prepared for review by referees before either a conference presentation or a paper submission, included with a published article or published separately in a collection of abstracts.

Likewise, to the extent that authors employing hypertext in their writing link directly from text citations to cited texts as suggested above, the bibliography (or reference list) can be treated as an independent text; a kind of author's suggestions for further related readings on the subject. One of the opportunities (or challenges, depending on your point of view) which electronic writing of this sort opens up is for more fulsome inclusion of author's notes.

Where it is not otherwise clear or evident how to represent these bonds in the text, hypertext authors might simply use the already ubiquitous "at" marker (@) together with the first letter of the particular bond to show its type (@A for abstract; @B for bibliography; @C for citations; @N for notes; @P for etc.)

These connections – footnotes, citations, reference lists, abstracts, reviews, commentaries, etc. – are part of the traditional "tricks" of the scholarly trade. They lend themselves well to adaptations to the electronic environment, simply by following up the appropriate logical implications with actual, physical connections. Collectively, they point directly toward the kind of docuverse proposed at the beginning of this paper.

Those who are troubled by the freedom ("loss of control") over the social science publications process suggested in this article should realize that the standard of publication introduced at the beginning is already in effect. There is nothing to prevent an enterprising author from creating an archive of manuscripts and papers and making these available to anyone who wishes to see them anywhere. This standard, of course, has been in effect since 1789, and we call it the First Amendment. What is different today is that the internet makes such publication practical, relatively easy (and most revolutionary of all) relatively low-cost. And, what is most lacking is a practical, effective system of peer-review.

Expanded Freedom of Ideas

At this same level, a number of new features might be incorporated as well. For example, implementation of a workable on-line peer review process might mean that rather than the now customary "accepted/rejected" status of manuscripts (which some in the field erroneously associated with "quality control") might be replaced by an expectation that the peer-reviewed status of a manuscript would be publicly evidenced by embedded links to the on-line comments of anonymous reviewers. In this way, faulty or inadequate research and argument could be publicly labeled and identifiable. Under the present system, "rejection" of manuscripts serves two undesirable purposes: It leads to the faulty conclusion that the quality of an individual's or group's research program is in direct proportion to the number of published articles it produces, rather than the theoretical or practical importance of what is

investigated. It also leads to something very much like censorship – the suppression of ideas.

Since the rejection of any particular manuscript is treated as a matter of private correspondence between the author, the journal editor and the reviewer(s), the audience must take largely on faith the integrity of the process. In some cases in the social science literature, that faith is seriously tested. Virtually everyone I've ever talked with on this issue can report tales of articles being rejected for highly dubious reasons. One of my personal favorites was a rejection I once received accompanied by a four-word review: "This article is wrong!" If that doesn't amount to censorship in the name of peer review, I'm not certain what to call it. The social sciences – which are paradoxically less committed to free and open scholarly communication than either the humanities or natural and physical sciences – might open up expanded horizons of freedom of ideas.

While the proposals offered in this memo may appear, at first glance to be impractical and utopian, they are in fact, highly practical and readily implementable. Those readers who remain unconvinced are encouraged to get access to Mosaic or some other WWW server and access the working model on line at the following URL: <expired link>

Copyrighting or Copy Writing?

The most complex and troublesome question posed by this proposal is the issue of copyrights. In social work, as in other fields, those who write usually have a primary interest in widespread dissemination of their ideas and expressions in the widest possible manner. Their primary interest in copyright law is in protecting claims of authorship, which may or may not involve major financial issues. (Total royalties on several book chapters I have published, for example, have never exceeded \$10.) At the same time, publishers are uniformly interested in protecting their investment – time and energy of editing, cost of paper and ink, printing technology, etc. – and in the case of commercial and "commercial nonprofit" publishers, making a profit. Such investments routinely run into the millions.

The category of "commercial nonprofit" is from Hansmann, 1990. It can be applied particularly to those university presses and other legally nonprofit publishers (like NASW Press and the historically important Association Press) who follow the lead of commercial publishers and skew their publications lists heavily in the direction of higher selling manuscripts, and do not pursue the traditional practice of "the book trade" (both commercial and nonprofit publishers) of also publishing what are considered "important" works with little sales potential.

It is ordinarily a matter of indifference to both reader and writer of social science literature whether the book market is up or down, paper costs are

rising or advertising competition is fierce. These are the unique concerns of publishers. It is, however, a matter of some considerable importance to a field with pretensions of an "applied science" whether the latest research reports confirm or discount existing practices. Yet, in the present environment, it is more the former than the latter which govern many aspects of the publications process.

The existing system of copy rights, while nominally intended to protect the integrity of authors, in fact, acts largely to protect the investment of commercial publishers. As print publishers have feared for several decades, the inherent nature of electronic hypermedia of the present is that they are quite easily copied and disseminated. Long term, this could lead to a parting of the ways of professional and academic writers and publishers.

The most radical proposal I wish to make is that social scientists begin immediately to adopt the following practices: That social science authors begin to assign copyrights for their own writings to the colleges, universities or social agencies which employ them. Then, it is a relatively simple matter for these institutions to grant reciprocal use privileges to one another. As soon as the practice becomes widespread, commercial publishers wishing to produce social science literature will have no choice but to agree to this. Any writer or publisher ... Such a system appears to be working quite well already in the case of software developed in educational settings.

In the particular context of present and emerging world-wide electronic internets hypertext and hypermedia technologies have important "self-help" implications which could enable social science to partially or totally by-pass the indifference and vagaries of the marketplace to the tasks of professional knowledge construction. In fact, if the examples from the original CREN on are studied, it can be readily seen how an electronic commons on the internet is likely to emerge in a more or less spontaneous, unplanned manner. It could be composed of the present editorial infrastructures of NASW, CSWE and the major independent journals, with connections to the doctoral, MSW, and BSW programs (and any social agencies or private practitioners with a computer and a modem). Much along the lines of voluntary board membership, each participant in this electronic commons would assume the costs of their own access ("transportation to the meeting" as it were) and agree to behave with an appropriate degree of decorum

We have already seen intimations of the potential for this through dramatic changes in the way research knowledge is produced and distributed: CD-Rom bibliographic data bases, a forthcoming CD-Rom edition of the **Encyclopedia of Social Work**, electronic co-authorship and review through the diffusion of "internet" addresses, list servers, gopher servers, etc. Such changes make it plausible already that distribution of the traditional "course bibliography" could be replaced by a set of keywords or search instructions.

They also point toward a future of electronically-distributed journals, including multi-media presentations mixing text, voice, video, graphics, etc.

The principal thesis of this paper, as noted, is that existing and anticipated hypermedia technology have already made possible the construction of an electronic web of all the related documents of social science (and any other academic discipline, professional field or interdiscipline). Actual construction of this web, referred to as the social science docuverse, together with exploration of the vast networks of linkages should be the major scholarly task embraced by social welfare scholars in the new century.

Thoughtful and intelligent construction of an electronic social science docuverse could result in radical transformations in the ways in which we cope with the information explosion, as well as ways in which the didactic portions of social science education are carried out. Education in policy, for example, might rely much more extensively on access to primary documents, like laws and court rulings, and upon commentaries on those primary documents created by faculty-scholars.

In a similar vein, direct practice might be taught through annotations of actual case histories with audio and video recordings and simulators guided by artificially intelligent "knowbots" able to simulate actual interview settings. In research, the data from key experiments, important surveys and other research projects could be made directly accessible to students for examination. Human behavior might even come to be taught entirely by observation and inference.

The most radical implication of turning such technology loose over the internet, however, is that it has the potential to make education truly space-independent. Existing and anticipated internet technology means that the social science docuverse need not be "housed" or located at any one site, but can be distributed across any number of institutions and campuses (including, conceivably, the entire range of accredited institutions). It also means that it can be accessible anywhere. This opens up possibilities of entirely new avenues of cooperation and competition between graduate and undergraduate programs. It also means that some of the existing antimonies between classroom and field could be "designed out" of the process of social science education, if we choose to do so.

References

Bush, V., Nyce, J. M., & Kahn, P. (1991). *From Memex to hypertext: Vannevar Bush and the mind's machine*. Boston: Academic Press.

Appendix A – Modules of a Docuverse

Text

1. A text is a set of sentences, each of which is a set of words. A text may thus be a sentence, paragraph, section, chapter, article, book, report, audio recording of human speech.
2. Any text may be the core object (referent) of a small and simple to large and very complex network of associated (collateral, subsidiary, and referential) texts and metadata. A full or partial set of related texts, together with their relations, can be defined as a hypertext.
3. A docuverse is a full set of related texts.

At a top level, a docuverse consists of a referent text and a set of linked, collateral texts, together with the subsidiary texts of each referent and collateral text in the set.

Referent Text. An original text of interest to someone for some reason. **(RT)**

Collateral Texts. Any user of any text (reader or writer) may link it with one or more other texts either symbolically or physically. One important way of establishing a collateral relation is creation of a referential text (RL), such as a footnote (FN) (embedded), reference list (RF) (attached) or bibliography (BI) (detached). **(CT)**

Subsidiary Texts:

Abstract. The third type of text is akin to a generally recognized abstract. It is a more complete description that may be short (150 words), medium (250 words), long (500 words) or extra long (1000 words). **(AB)**

Analect. The fifth type of statement is an extended passage (up to 5,000 words) from a longer work. **(AN)**

Annotated Bibliography. A bibliography (BI) combining citations (CT) and annotations (AO). Ordinarily, the most common annotation added to a bibliography will be an abstract (AB). **(AY)**.

Annotations. Explanatory or critical comments added to a text by someone other than the original author(s). **(AO)**

Brief. The second type of statement attempts to briefly (50 words or less) describe the contents of a text in most general terms. **(BR)**

Bibliography. A free-standing referential text (RL) composed of two or more citations (CT) and detached from any of the Referent Texts (RT) it links to. **(BI)**

Critique. A written, audio or video assessment or set of opinions about a referent text. **(CR)**

Citation. An element of a bibliography (BI), footnote (FN) or other referential text (RL), consisting of identifiers (ID) and locators (LO), and in some instances descriptors (DE).

Descriptor. The first type of hypertext text-descriptor is one short phrase. It is externally focused and attempts to place the work by type (e.g., theory, survey, case study, etc.) and place it within a content area (e.g., nonprofit studies, urban studies, etc.). **(DE)**

Digest. An abstract, outline, synopsis, or summary of a referent text prepared for the express purpose of bringing a different order or sequencing to the points made in the original text. **(DI)**

Footnote. An embedded referential text (RL) included within a referent text (RT).

Identifier. (ID)

Keyword. A single word or phrase used to accurately describe or characterize a referent text. **(KW)**

Keyword In Context (KWIC)

Locator. (LO)

Outline. A detailed, sequential list of the main points of a referent text hierarchically organized. **(OU)**

Précis. A shortened version of a referent text containing the main points and omitting examples, arguments, and other details. **(PR)**

Reading. An oral repetition, rendering or interpretation of a text. **(RD)**

Review. A text offering an assessment or set of opinions about a referent text especially one following a pattern or form. (E.g., book review, movie review, etc.) **(RV)**

Summary. A brief, abstract, précis, outline, digest or synopsis of a referent text that purports or aspires to a thorough or complete recapitulation of the original text. **(SU)**

Synopsis. A condensed version of a referent text. **(SY)**

Referential Texts:

Links. Two or more texts may be linked (i.e., they can point to one another) in many ways. Two important ways are:

Hypertext link (HL). An electronic pathway that, when activated, automatically gives the reader access to a second text.

Syntactic link (SL). Traditional, pre-electronic links between two or more texts including footnotes, reference lists, bibliographies.

Syntactic links generally consist of one or more of three components:

Identifiers – Symbols that uniquely define the text

Locators – Symbols that aid in finding and securing texts.

Descriptors – Symbols that describe or characterize a referent text.

Links – Pointers to and from other texts.

Lohmann, 1992.

A theory of nonprofit activity and structure. (Lohmann, 1992.)

A theory of nonprofit activity and structure premised on the centrality of the concept of a commons characterized by voluntary participation, shared resources, shared purposes, and emergent values of mutuality and justice.

An examination of the effects of planning and zoning laws on cities. (Sennett, 1972)

Argues that urban planning and zoning laws produce excessively ordered cities with segregated, stultified cultures. (Sennett, 1972)