11-29-2001

Knowledge Management Technology: Will There Be A Second Chance?

Roger A. Lohmann
West Virginia University, roger.lohmann@mail.wvu.edu

Knowledge Management Technology: Will There Be A Second Chance?

Roger Lohmann, Ph.D.
West Virginia University

Introduction

There is a level of *reflexivity* involved in all matters of ARNOVA member concerns, in that many of the same issues and concerns which draw the attention by ARNOVA members as matters of research, theory and practice are also useful for understanding and elucidating the nature of membership in ARNOVA itself. This is true in a manner that does not apply also to, say a physical science research organization or a historical society.

For example, it is possible to understand ARNOVA fairly narrowly as a *membership association*, with a voluntary, paid membership, nonprofit corporate status, an elected board responsible for managing the affairs of the organization, and a small paid staff. Although this membership resides predominantly in the United States and North America, international membership from more than 40 other countries is an increasingly important dimension of the association and its activities.

However, it is also possible to see ARNOVA as an organization defined by time. As a temporal organization, ARNOVA has evolved from its earlier incarnation as AVAS (the Association of Voluntary Action Scholars) which was largely an episodic, or *Cinderella* organization which literally came into existence once a year around the annual conference and relapsed to a largely dormant state for most members much of the rest of the year until the next annual conference. (Lohmann, 19XX) Currently, ARNOVA has a much more continuous, full-time program of activities, committee work and membership interaction through such media as the electronic discussion lists. This is not simply a matter of having a full-time paid executive and staff. Many members, too, are involved on a year-round basis.

As a cultural organization, the year-round ARNOVA has emerged as one of the major contributors to the evolving and increasingly clear paradigm which goes by various labels, including nonprofit organizational studies, third sector studies, civil society

This paper was presented at the 2001 Annual Conference of the Association for Research on Nonprofit Organizations and Voluntary Action (ARNOVA), Miami, FL. November 29, 2001.

studies, nonprofit organizations, voluntary action and philanthropy studies and the like. Also, from its earliest days of interdisciplinary membership primarily from 2-3 major disciplines, the association has grown to encompass membership from nearly three dozen disciplines.

It is also possible to understand ARNOVA somewhat more broadly as a temporal, cultural and infrastructure organization. Although it remains for one or more formal presentations of a full-blown analysis or identification of this paradigm

A principal concern of this paper is the role of the association, through its board of directors and its membership, in proactive management of the cultural organization of third sector studies.

In looking at this issue, use will be made of the concept of esoteric concept called infomatics and the currently rather mundane informatic applications of *knowledge management*. Infomatics is a very useful term that is in fairly widespread use in the British health care system, and meaning roughly *computing and digital information processing*. Knowledge, of course, is something we are all involved in every day. Research is concerned with the production of knowledge, publication with the dissemination of knowledge, teaching with the interpersonal and intergenerational transfer of knowledge, and policy and practice with the application of existing knowledge to practical problem solving. At each of these points in what can be highly complex interplay between neural, linguistic, cultural and infomatic networks, knowledge – or rather the artifacts of knowledge – may involve anything from conversations and other live verbal performances, written documents, audio and video recordings, digital files of all sorts, and all the other detritus of contemporary knowledge transfer. In this paper, my principal concern is with the present and possible uses of infomatic networks to facilitate and enhance the management of knowledge of the third sector by ARNOVA members and third sector researchers more generally.

While this issue might be approached broadly through the traditional philosophical lenses of epistemology and ontology, dealing with knowledge in a philosophical sense is a notoriously difficult, abstract and arcane pastime of interest only to a limited few. The concern here is a much more immediate and practical one that might be thought of as the role of the association in the production, dissemination and application of knowledge about the third sector the flow of scholarly documents generated by ARNOVA members primarily for the use of other members.

This is not a new interest for me. I wrote a number of years ago of the ARNOVA *docuverse*, or the universe of paper and electronic texts that future historians or archeologists might consider the principal artifacts in which the paradigm of third sector studies is grounded. (Lohmann, 1996) This docuverse can be thought of as potentially encompassing all of the past, present and future scholarly papers, presentation outlines, handouts, overheads, data sets, published articles, books and chapters, and other artifacts *as well as all of the intentional (logical, referential and historical) linkages between them.*

This was not a particularly novel idea then or now. A few years earlier, in 1991, Stephen Harnad set forth one of the earliest and clearest statement of an electronic docuverse for physical sciences in the following statement: "The American Physical Society's Task Force's Report on Electronic Information Systems … has sounded all the right chords: The idea is to develop a world scientific information system that will include all the formal scientific literature that has been, is being, and will be published, as well as the informal unpublished scientific communications that surround it…" (Harnad, 1991) It is part of the conventional wisdom of computer networking that even earlier – in the 1940s – Vannevar Bush first proposed his Memex system and the original creation of HTML by Tim Berners Lee was directed at much the same insight among physicists as Harnad refers to.

The challenge today, as it was then, is how to go about structuring such a system in order to facilitate, rather than impede what Thomas Kuhn taught us to call "normal science." There have been many different approaches to this idea of furthering the docuverse over the past decade. Harnad went on to write that documents in such a system would be "…all in an electronic form that is searchable and accessible by any scientist anywhere in the world. The system would be furnished and administered by a collaborative effort among the scientists themselves through their learned societies, universities and libraries, and perhaps some publishers and data-base producers too." (Harnad, 1991)

Almost accidentally, due to the reality that almost all of us use word processors and most of us are linked to the Internet through office LANs, the first criterion of creation a galaxy of electronic documents potentially accessible by "any scientist anywhere in the world" has almost certainly already been met. In fact, anyone who looks closely at the issue will quickly realize that actually making the appropriate contents of our word processing files accessible to the

entire body of nonprofit researchers in the world is not a particularly challenging problem. The primary missing ingredients at the present time are the necessary institutional commitments from our learned society – ARNOVA – and our universities, libraries, publishers and database producers. And it may, in fact, be some time before we see that develop as each of these in their own way listens to the siren songs of relative advantage and potentially huge profits that probably will never materialize.

To begin consideration of this question of knowledge management, we might ask about the basic purpose of ARNOVA and what ARNOVA members expect to get from participation in the association. From a knowledge management standpoint, this matter resolves to a single primary question: Why do scholars and scientists publish? Although there are no doubt materialists whose some interest is in the impact of publishing on incomes, and careerists whose primary interest is to enhance their resumes for professional advancement or perhaps even to market their words, surely the agreed upon motive of the true scholar/scientist is to advance human inquiry. And, just as surely, such an enterprise is and always has been a collective, cumulative and collaborative one. From this vantage point, scholars publish in order to inform their peers of their findings and, equally important, to BE informed by them in turn, to INTERACT with them, in the cycles of reciprocal influence that constitute an evolving body of scholarly research. In a word, the purpose of scholarly publication is COMMUNICATION -- with peers, and for posterity." (Harnad, 1991)

From this perspective, Harnad argues, "electronic publication is not just a more efficient implementation of paper publishing, it offers the possibility of a phase transition in the form (and hence also the content) of scholarly inquiry." This "phase transition" has at least two principal implications: First, it becomes clear that scholarly inquiry has always been a continuum of documents, stretching from the earliest and sketchiest proposal through various finished products. Secondly, the boundaries between "informal" scholarly literature like research memoranda and conference papers and "formal" literature like book chapters and journal articles are blurred. (Harnad, 1991)

It is widely agreed that peer review is the distinctive editorial characteristic of the system of scholarly publishing. "Peer review (Harnad 1982, 1985, 1986) is often cited as the boundary between the informal and the formal literature, but peer review too is only a matter of degree. The "prestige hierarchy" among journals corresponds roughly with the level of rigor of their refereeing

systems, and this is probably the true function of peer review: To serve as a quality filter according to which one can adjust how selectively one reads the literature relevant to one's interests and expertise. Another function of peer review is of course to give the author critical feedback in helping him report (and do) his work correctly and clearly. Both these refereeing functions can be implemented in the electronic medium too -- for both primary articles and comments -- and much more quickly and efficiently than in the paper medium (as noted in the APS Report)." (Harnad, 1991)

Peer review has also often been cited as a barrier to informatic knowledge management in this area, but Harnad has been one of the most vocal and public critics of this idea. "I have described elsewhere some proposals for implementing peer review on the net in the form of hierarchies of peer-reviewed groups, with read/write access for the peers in a given specialty or subspecialty at level (as in an academy of science) and read-only access for everyone else at level i-1 (but with the possibility of posting at one level higher through a read-write peer at that level -- the equivalent of an editorial board member in the present paper system), extending all the way down to an unrefereed vanity press at the bottom." (Harnad, 1991)

Six Components

In the second part of this paper, I would like to present and explore briefly six components of an overall solution to this problem of the docuverse. They vary from the simple to the slightly arcane. They include: open standards; peer-to-peer networks; a semantic web; cover-page semantics; resource description frameworks; and XML embedded documents.

Open Standards

One of the major developments in the future of ARNOVA, all university-based members and their universities ought to be a robust commitment to Open Source/Open Standards. As an example, the schematic for an XML-DTD offered in Appendix C of this paper is set forth as such a possible open standard and the DTD for the Rich Site Summary (RSS) has already achieved that status.

Peer to Peer Networks

Adrian J. Pullen, Senior Lecturer at the Northeast Wales Institute of Higher Education, defines Peer to Peer (P2P) Networks as a small group of computers connected together, sharing resources with no central computer and no centralized resources, such that any computer can share resources with any other. (www.newi.ac.uk/pullina/peer/sld002.htm) Pullen argues that there are a number of reasons to use Peer-to-Peer Networks including low cost, the use of existing equipment, ease of setting up, ease of operation, and ease of expansion.

In the wake of the publicity surrounding Napster in recent years, most people should recognize the underlying idea of P2P, as well as the centrality of the copyright question. They may, however, also associate the idea with piracy, hackers, or a number of other unsavory ideas.

In essence, a group of users grant limited access to some of the files on their computer that are gathered together in a single location such as a folder or sub-directory in exchange for similar access to the similar files of others. Using an elementary technology like FTP (file transfer protocol) or several other available means, establishing basic P2P networks for sharing documents is a snap. When combined with other widely available technologies, like the ability to print Word documents to disk as Adobe Acrobat .PDF files, a docuverse auxiliary to the existing conferences-and-journals dissemination system begins to come into view.

Peer review is one of the fundamental parameters of scientific communication, so it is probably essential that any docuverse system operate on a peer-to-peer basis.

Cover Pages

Simple transmission of documents is not the only important issue, however. There is also the question of readily identifying, placing and classifying document contents. One such not-very-interesting solution would be to rely on the unaided facilities of Internet search engines. (Another would be to circulate RSS links of the type I referred to in my ARNOVA paper last year. See discussion below.)

Bipin C. Desai of the Department of Computer Science at Concordia University in Montreal, Canada, laid out an (undated) online proposal for what he calls cover page semantics. A cover page (or semantic header), he says, "is a portion of each document which should contain information useful in searching for a document based on a number of commonly used criteria." This is, in essence, a more expanded version of the Meta-tag already in common use in HTML based web pages.

Why would one want to use such cover pages, you might well ask. Desai's answer is "The information from the semantic header could be used by various indexing schemes to help locate appropriate documents with minimum effort."

Once the basic concept is grasped, it is easy to envision, "that regional and/or specialized databases would be created to maintain archives of cover pages. These databases could be searched by cooperating distributed expert systems to help users locate pertinent documents." As a straightforward extension of its current functions, for example, the ARNOVA web site might easily contain a searchable archive of such cover pages along with suitable addresses and other information for gaining FTP access to desired documents.

In his online discussion, Desai proceeds to develop this idea along the lines of a library or general information archive to outline his schema for such a system.
(www.cs.concordia.ca/~faculty/bcdesai/cindi-system-1.0.html)
Although some of the additional details are interesting and worth further consideration, we can take two main points of general use away from this consideration:

1. The use of SGML/HTML/XML type tags (with content delimited within matching opening and /closing tags)
2. The description of this type of activity as a "semantics" of knowledge.

A Semantic Web

This idea of a semantic approach is developed further by Tim Berners Lee, the original author of HTML and god-father, as it were, of the World Wide Web. In a September, 1998 on-line paper for the World Wide Web consortium Lee outlines his plan for a semantic web, which he defines as "a set of connected applications for data on the Web in such a way as to form a consistent logical web of data". (www.w3.org/DesignIssues/Semantic.html)

Such a semantic web, he notes, "is a web of data, in some ways like a global database." This, then, is another general insight of use in the development of a docuverse:

A network of linked documents of the type suggested by Bush, Lee, Harnad and others would have some of the characteristics of a large, rather unwieldy, database in which texts (abstracts, manuscripts, conference papers, book chapters, reviews, research memoranda, et. al. constituted the records. Thus, the docuverse should be in essence a peer-to-peer database. This insight leads directly to another.

The Resource Description Framework (RDF)

According to Lee, "When looking at a possible formulation of a universal web of semantic assertions, the principle of minimalist design requires that it be based on a common model of great generality. Only when the common model is general can any prospective application be mapped onto the model. (That) general model" says Lee "is the Resource Description Framework." (www.w3.org/TR/PR-rdf-syntax/)
The RDF specification, which was still in draft form as a possible Internet standard as of this writing, is "is a model for representing named properties and property values." Lee's general approach to a semantic web begins with what he calls "a basic assertion model", consisting of two primary resources: assertions and quotations (which are assertions about assertions). It is built upon three principal elements: *resources*, or the things being described in RDF expressions; *properties*, or specific aspects, characteristics, attributes, or relations used to describe resources; and *statements*, made up of property statements associated with particular resources. According to the preliminary specification, "These three individual parts of a statement are called, respectively, the *subject*, the *predicate*, and the *object*." (Hence the reference to semantics.) He says also "The semantic web data model is very directly connected with the model of relational databases." (www.w3.org/DesignIssues/RDFnot.html)

In my 2000 ARNOVA paper, I gave a brief explanation of a particular RDF already in widespread use.[1] It is called the RSS (or, Rich Site Summary), which is reproduced here:

```
<item>
<title>RSS Resources</title>
<link>http://www.webreference.com/authoring/languages/xml/rss/</link>
<description>Defined in XML, the Rich Site Summary (RSS) format has
quietly become a dominant format for distributing headlines on the Web.
Our list of links gives you the tools, tips and tutorials you need to get
started using RSS. 0323</description>
</item>
```

An RSS consists of a pair of tags delimiting the container, a title, a web address link and a brief description. As standards go, it is an extremely simple one and yet a powerfully concise summary of location information. Any program written to process documents in a system of peer-to-peer networked databases would be able to identify

[1] These comments were written several years before RSS became a widespread feature of web usage.

any particular document/record in the database from this information. A listing of such listings would, perhaps, be most analogous to a bibliography.

Thus, for example, an item distributed to the ARNOVA-L discussion list and written in RSS could, at any point, be picked up from the archive and processed What would such an item consist of? Here is an RSS container for the 2000 ARNOVA paper and the accompanying PowerPoint presentation.

<item>
<title>Furthering the Scholarly Commons</title>
<link>http://www.wvu.edu/~socialwk/faculty/Rlohmann/fsc.htm</link>
<description>
Paper by Roger A. Lohmann, Professor of Social Work. West Virginia University. Presented at the 2000 annual conference of the Association for Research on Nonprofit Organizations and Voluntary Action, New Orleans LA.
</description>
<link>
http://nova.as.wvu.edu:8080/nova/Members/rlohmann/Manuscripts/FurtheringTheScholarl%233F2A1.pdf/file_view</link>
<description>
PowerPoint presentation accompanying paper entitled Furthering The Scholarly Commons by Roger A. Lohmann, Professor of Social Work, West Virginia University.
</item>

From all of this, we can also take away two relatively simple points:

1. It is becoming increasingly possible to visualize knowledge management in the social sciences as the construction of semantic webs, linkages and connections between meanings and not simply formalisms. In order for such meanings to be clarified, however, there is a strong role implied for authorial intent.
2. Regardless of its present form or formats, it is possible to visual the documents of a scholarly association like ARNOVA (or, still more broadly, "the third sector literature" such as conference papers, journal articles and books, and such fugitive documents as research notes, datasets, memoranda and reports) as records in a vast and complex virtual database to which members contribute as peers.

At this point, one main body of computer science interest in this topic moves over into artificial intelligence programming. Lee's

paper, for example, explicitly references the CYC® Inference Engine (www.cyc.com/tech.html#cycl)[2], and its accompanying representational language, and the Knowledge Interchange Format (KIF). It would, at this point, require the widespread dissemination of specific products, analogous perhaps to SPSS or Word for this tangent to hold much interest for the main body of ARNOVA members.

[2] This link is no longer active.

XML-Embedded Documents

A fourth development of some interest in this regard is the emergence and widespread adoption of XML (Extensible Markup Language). XML may be thought of as a step-sister of HTML and child of SGML. With its Resource Description Framework (RDF)-style assumptions, it holds considerable potential for providing real meaning to the idea of semantic nets. In particular, creation of entirely new tags and its potential for creative organization of systems of tags into Data Type Definitions (or DTD's), offers great potential as the structural basis of such the virtual database of a field of knowledge.

Thus, as the example in Appendix 3 shows, it would be possible (in fact, fairly straight-forward) to write a standard DTD that names all of the elements in a standard research report. If these tags were coded as hidden text in an ordinary word processing document and available as part of a database of documents in a peer-to-peer network, it should be possible for any member of the system to, for example, quickly and easily gather a census (full sample) of sample sizes used in studies of board governance research, or pull together a representative list of the findings from a sub-sample of studies.

Conclusion

When we put all of these pieces together – P2P networks composed of documents with their own cover pages containing appropriate meta-text statements and trailing bibliographies including RSS tags, linked into semantic nets defined within the bounds of content-appropriate Resouce Description Frameworks and XML-based Data Type Definitions and where each peer member has both originate and retrieval capabilities – we have gone a very far ways toward the creation of a docuverse and a scientific commons like the ones suggested by Vannover Bush, Tim Burners Lee, Stephen Harnad. Moreover, such a basis for knowledge management would be based not on the pecuniary principles of the marketplace or the centralized authority of the state but on the mutuality of the commons and the open standards movement.

Appendix A
A Sample Cover Page

<semhdr>

* <title> ..... </title>
* <subtitle> ..... OPTIONAL </subtitle>
* <alttitle> ..... OPTIONAL </alttitle>
* <char-set> ..... OPTIONAL </char-set>
* <Language> ..... </Language>
* <author>
* <aname> ..... </aname>
* <aorg> ..... </aorg>
* <aaddress> ..... </aaddress>
* <aphone> ..... </aphone>
* <afax> ..... </afax>
* <aemail> ..... </aemail>
</author>
* <Subject>
* <General> ..... </General>
* <Sublevel1> ..... OPTIONAL </Sublevel1>
* <Sublevel2> ..... OPTIONAL </Sublevel2>
</Subject>
* <Keyword>
* .....
</Keyword>
* <Dates>
* <Creatred> ..... </Creatred>
* <Expiry> ..... </Expiry>
* <Updated> ..... </Updated>
</Dates>
* <Version> ..... </Version>
* <Hardware>
* .....
</Hardware>
* <Software>
* .....
</Software>
* <Coverage> ..... </Coverage>
* <Classification> ..... </Classification>

```
*       <Annotation> ..... OPTIONAL </Annotation>
*       <URL> ..... </URL>
*       <URN> ..... </URN>
*       <UAS> ..... </UAS>
*       <Cost> ..... </Cost>
*       <abstract> ..... OPTIONAL </abstract>
*       <size> ..... </size>
</semhdr>
```

Appendix B
A Sample DTD for a Social Science Research Report

<ABSTRACT>

This tag is intended to denote authorial or publication produced abstracts describing the contents of a text.

<ARGUMENT>

An optional tag used to identify succinct statements of the position or point of view of a text and included within the text itself. (Different from an abstract). This tag may be used repeatedly and any one of them may contain a nested structure of an indefinite number of sub-arguments. (May be replaced by the SHOE DTD).

<CITATION>

The inclusion of identifier-type information within the body of a document.

<COMMENT>

A one or more sentence brief statement in one text about another. One of the basic document types, along with Abstract, Précis, Paper, Article, Chapter and Book.

<CONCLUSION>

Any inference(s) drawn by the author(s) of a text about its broader meaning or implications, whether based on the logic or an argument, the presentation of one or more findings, or the simple statement of a position.

<FINDING>

This tag is used to denote statements or determinations based upon consideration of a set of evidence.

<HYPOTHESIS>

This tag is used to identify explicit statements of the hypotheses under study or tested in a research investigation.

<INDEX>

This tag is used to denote a manual or page number index of a document.

<MEASURE>

This tag is used to denote discussions of scales, questionnaires, instruments and other paraphernalia used to construct or generate quantitative, numerical or other measurements.

<QUOTE>

This tag is intended to be used only for direct quotations.

<REFERENCE>

This tag can be used for all types of indirect quotations and references to other works.

<SAMPLE>

This tag is used to denote discussions of a sample, as well as discussion of how the sample was drawn, its limitations, and other related characteristics.

<SCALE>

This tag will usually be used in conjunction with the measures tag (either within a measure tag or surrounding one)

<STUDY TITLE>

This tag may be used if the study from which a particular text is derived is identified separately from any specific publication arising from it.

<TITLE>

This tag is used to identify the title (or similar pneumonic identifier) of a document.

<TOC>

This tag is used to denote the Table of Contents of a publication. It may be used, alternately, to construct an outline, listing of sub-headings or other indications of the structure of a document and placement of items within it.

<UNIT OF ANALYSIS>

This tag may is intended to denote explicit identifications of the unit of analysis of a study.

Appendix C

SHOE (Simple Hypertext Ontology Extensions)

One interesting intermediate set of developments illustrating this trend is SHOE or the Simple, hypertext ontology extensions, created by the Parallel Understanding Systems Group in the Department of Computer Science at the University of Maryland at College Park. SHOE is described as "a small extension to HTML which allows web page authors to annotate their web documents with machine-readable knowledge." (www.cs.umd.edu/projects/plus/SHOE/index.html)

SHOE adds the following tags to the HTML standard:

To make possible ontology declarations, SHOE adds

`ONTOLOGY, /ONTOLOGY,`
`USE-ONTOLOGY,`
`DEF-CATEGORY,`
`DEF-RELATION, /DEF-RELATION,`
`DEF-ARG,`
`DEF-RENAME,`
`DEF-CONSTANT,`
`DEF-TYPE,`
`DEF-INFERENCE, /DEF-INFERENCE,`
`INF-IF, /INF-IF,`
`INF-THEN, /INF-THEN,`
`COMPARISON, /COMPARISON,`
`CATEGORY,`
`RELATION, /RELATION,` and
`ARG.`

To make possible semantic markup of HTML pages, SHOE reuses some of the tags from above and adds `INSTANCE` and `/INSTANCE`. The other tags used for markup are `USE-ONTOLOGY, CATEGORY,`

RELATION, /RELATION, and ARG. Additionally, SHOE declares the META HTTP-EQUIV tag "SHOE".

SHOE adds the following tags to the HTML standard:

To make possible ontology declarations, SHOE adds ONTOLOGY, /ONTOLOGY, USE-ONTOLOGY, DEF-CATEGORY, DEF-RELATION, /DEF-RELATION, DEF-ARG, DEF-RENAME, DEF-CONSTANT, DEF-TYPE, DEF-INFERENCE, /DEF-INFERENCE, INF-IF, /INF-IF, INF-THEN, /INF-THEN, COMPARISON, /COMPARISON, CATEGORY, RELATION, /RELATION, and ARG.

To make possible semantic markup of HTML pages, SHOE reuses some of the tags from above and adds INSTANCE and /INSTANCE. The other tags used for markup are USE-ONTOLOGY, CATEGORY, RELATION, /RELATION, and ARG. Additionally, SHOE declares the META HTTP-EQUIV tag "SHOE".

Harnad, S. (1992) Interactive Publication: Extending American Physical Society's Discipline-Specific Model for Electronic Publishing. Serials Review, Special Issue on Economics Models for Electronic Publishing, pp. 58 - 61. (www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad92.interactivpub.html (Harnad, 1992 #1))

www.arl.org/scomm/subversive/sub02.html (Harnad #2)