

# *The Ethics of the Click: Users and Digital Information in the Internet Age*

NANCY R. JOHN

University of Illinois at Chicago, Chicago, Illinois, USA

---

More and more information seekers are turning to the World Wide Web as a source of ready and authoritative information. Many of these Internet users do not fully understand the new information environment in which they seek information and how it differs from the print information environment. Lured by the easy click, seduced by colourful Web sites, and convinced that the number of items they can

find through Internet search engines (e.g., AltaVista, HotBot etc.), these information seekers believe that up-to-date quality information is free and only a dial tone and a click away. This paper examines the changing information environment from the standpoint of information providers and looks at the hidden agendas and values that conflict with the concept of cheap, easy-to-get, comprehensive, accurate information.

---

## *Introduction*

Despite some estimates of more than 300 million Web pages and terabytes of Internet-accessible data (for actual figures of hosts for World Wide Web see Zakon 2000), there remains a shortage of access to high-quality electronic information. The conversion of paper-based information to electronic Web-accessible information has increased at a staggering rate during the past two years; the amount of information being born digitally has also increased enormously. And still, information becomes available via Internet in incredibly haphazard and non-systematic ways. The result is no less than a crisis for those professionals who wish to provide access to high-quality information in a systematic way and for those who wish to make informed decisions in the timeframe expected by today's consumers. This paper explores how this situation has come to be using an encyclopaedia and some medical literature as case studies, and what hopes there are for its amelioration in the future.

## *Information before the Web*

In the beginning, not so long ago, in the 1950s, there was chaos. Access to the research literature of medicine was by-word-of-mouth or by glancing through the journals in one's inbox.

But the literature of medicine was beginning one of the sharpest growth trends this century would see, as a handful of quality medical journals would turn into hundreds and then thousands of journals. In the United States, researchers at the National Library of Medicine would begin the indexing and documentation of 771,287 items published from 1960 to 1965 using original Medical Subject Headings as keywords in the Index Medicus (NLM 2000).

Beginning in 1966, the Medlars Online project would not only provide access to medical literature, but would become one the world's premiere bibliographic databases for the next four decades. MEDLINE includes more than 9.2 million records from 3,900 current biomedical journals published in the U.S. and 70 other countries. It is growing at

Nancy R. John is Associate Professor, Assistant University Librarian and Manager of Library Systems at the University of Illinois at Chicago Library, Box 8198 m/c 234, Chicago, IL 60680 USA. Tel: +1 312 996 2716 Fax: +1 312 413 0424. E-mail: nrj@uic.edu

This paper was presented at the conference "Scientific Publishing Under Change" organised by the Faculty of Health Sciences, Aarhus University and the State & University Library, Aarhus, Denmark, to mark the opening of a new medical library and education centre, January 2000.

more than 31,000 citations per month. The MeSH heading searching and the full citations with abstracts were the cutting edge indexing technology B.T.W. – before the Web.

### *Electronic information strategy*

But like all of us in the information business, they have been forced to compete with the likes of Yahoo, Altavista, and amazon.com. But unlike Yahoo and Altavista who only indexed things born electronically, the great medical indexing service has had to maintain its full coverage in a world that is not yet 100% full-text. And perhaps more importantly, they do this in a world where standards for full-text have not yet emerged nor is there a widespread understanding of what these standards might be and there is not yet agreement on which marketplace forces effecting our information economy are important.

There are three fundamental forces that limit access to electronic information today. These are:

- How much older, print-based information has been converted to digital format;
- How much new information is coming out in digital format;
- Whether the electronic information is network accessible.

The first category, the lack of data conversion, usually can be traced to a perception of the actual value of information in the marketplace. Since the information is usually available in one format (generally print), it must be shown that there is some exceptional added value to re-issuing it in another (digital) format. The conditions pertaining to the issuance of the original print item can affect this greatly. For example, an item that is only available as a print item will require the capture of the text into a digital format. An item that was printed from an electronic manuscript requires very little effort if the goal is simply electronic display. If, however, the goal is make an *electronic* item – to enhance access by adding navigation features such as hyperlinks – the costs of the electronic edition can increase substantially.

The second category, how much new digital data is being issued, is difficult to analyse succinctly. Marketplace considerations are very real

and very complex. The lack of the equipment to use electronic formats will reduce the number of potential customers for digital information. Thus we see some electronic titles attaining a ubiquitous presence – Microsoft's Encarta Encyclopedia – because they are offered bundled with equipment purchases. Simply put, it just doesn't pay to digitise information that few can use, and even fewer will pay for.

The final area, whether electronic information can be used over the network, presents a different set of considerations. Traditional data-processing professionals may not be knowledgeable about internetworking issues or understand the intrinsic value of information. In some cases, data-processing staff can be knowledgeable about the audiences for electronic information, but are extremely concerned about the security risks that open networks and information access present; financial experts are concerned about the economic risks of open access. Newspapers provide a good example of the financial dilemma; today's "paper" paper is free on the net but costs money at the newsstand; yesterday's paper is free from the trashcan, but articles cost \$2.95 from the net.

In both these cases, what is needed is the ability to convert data from a proprietary format to an open format and to provide a low-risk access method. Librarians bring their desire to open up access to information and have pushed network administrators to invent ways to make information more widely available in secure ways. At the same time, in the case of journals, libraries are maintaining our paper subscriptions to assure long-term access for our users and to provide economic stability for the publishers to experiment with pricing models.

### *Enter the Web*

Into this somewhat conservative and gentlemanly environment enters the World Wide Web. Immediately, the standard for access becomes immediate. There is no expedite in a world full of hyperlinks and clicks – instead there are slow links and fast links. And the land of no links is a place where – well, it's not where you want to be. Furthermore this high-speed chase for the hottest, newest thing has pushed the information economy into a world that uses new rules, if in fact there are rules at all. Some argue that it's at-

tention that drives this economy where too much information is competing for your attention. But whatever it is, we can all agree there is too much stuff. A sociologist at Yale begs libraries to keep information from him – he says that information seeks him everywhere in this world of email, fax and telephone. He tells of being thankful his professional journals come in plastic sleeves – he only reads an article when he can recall three different sources pointing it out. As a result every six months he can throw out most of his professional reading still in its plastic sleeves.

Enter the click. What are respectable indexes and respected libraries to do if they are to continue to serve their users?

### *Will Printing Disappear?*

“... no creditable model for the Web replacing print” – Frank Romano (1998)

Books seem unlikely to migrate quickly except for those that are constrained by their linearity. There is some evidence (e.g. Ingram) of movement toward on-demand publishing for books that don't have a large audience. Ebooks, while a curiosity and a lot of fun, do not seem to be overtaking the mass market. Romano (1998) claims this is because of the three B's: bedroom, beach and bathroom.

The journal marketplace is somewhat different: specialised journals and trade publications have a strong paper future; mass-market journals have a very strong future; academic journals whose economics in a paper world was problematic are going electronic at a staggering rate.

Maybe there is not creditable model, but a lot of publishers are trying to be the one who discovers the best approach. And despite this prediction, a lot of people think that clicking will get them good, cheap, information. But when no one understands the rules of the click, some interesting things can happen.

### *Encyclopaedia Britannica*

How the Encyclopaedia Britannica tradition adapted to the Web is a lesson in the power of the click. The Britannica name had prestige and status. The Encyclopaedia had a long history. It was known for the authenticity and the accuracy

of its content. Some said it was a bit stodgy. Parents agreed that while it was expensive, it gave their children an important advantage in studying and researching papers for school.

The management at Britannica failed to perceive the true threat of electronic publishing, thinking that their history would stand them in good stead. They were slow to recognise that they were being displaced by home personal computer purchases. Further, encyclopaedias, edited and compiled with different editorial values were coming bundled with these personal computers. Consumers were accepting the trade-off of spending their “encyclopaedia” dollars for another tool that would give their children an edge and prepare them for the future.

When the leadership of Britannica realised that they had to go online to survive. To their credit, they created an ambitious programme to maximise the digital opportunities and advantages. This included recreating the look and feel of Britannica without the clunkiness of the paper index and difficult page navigation of the paper Encyclopaedia Britannica. They also developed better searching (and a vastly improved index). When they realised that page counts and volumes weren't an economically imposed restriction, they expanded the coverage of topics. They also saw that (lack of) timeliness no longer needed to be an issue. And finally that there were lots of ways to enrich the content: some glitzy technical features weren't necessarily bad and the online version could interact with the print and CD-ROM version.

The Encyclopaedia Britannica offers Britannica Online by subscription over the Internet's World Wide Web. This version has not only the 44 million words of the original print edition but also includes pointers or hyperlinks to Internet-accessible information such as the budget of the United States or the Vatican exhibit at the Library of Congress. Then they discovered that the content could be unbundled and repackaged and it is now being used to develop new theme resources – such as a site on Nobel Prize winners or on the Olympics. In the latter case, the historic 200-year old encyclopaedia is cashing in on the current interests of its users – recently it debuted a site of the Titanic. But subscription and sales of spin-off products turns out not to be the only financial model, so britannica.com (a free site sponsored

by advertising) premiered last September. Which is a better model? Time will tell.

### *Other examples*

We could find several other new models for publication and distribution of printed matter via the Web. Some magazines are giving the Web version away, paying for it by advertisements and by the increased sales of paper that free Web access promote. Others, like the *Chronicle of Higher Education*, provide subscribers to the paper version with daily news alerts via email and special pointers to upcoming paper editions. Some newspapers, like New York Times, may give away access to today's paper to some readers (in the U.S.), but they charge for access to older information. Some bookshops have gone online; some booksellers (Amazon.com) are only online; and some booksellers actually don't sell books but instead point you to others those who do in a more effective way (by searching across booksellers, allowing comparison of out of print prices, etc.). So many models are available to us. So many models are available to us, that it is not only confusing to us, it confuses the information providers when they try to respond to these different models.

### *American Pediatrics Association*

Let's look at the case of MEDLINE and a journal called *Pediatrics* – a highly respected publication of the American Pediatrics Association. The decisions these two separate entities made became hopelessly intertwined as reasonable people search for reasonable responses to a changing marketplace.

The National Library of Medicine has licensed MEDLINE over the years – once for a high enough price to fund the quick turnaround time in its indexing, and now for a much lower price because some U.S. citizens successfully argued that U.S. taxpayers were double-paying for MEDLINE's development. NLM felt it was in the business of making the content for the index and for providing professionals with a way to search it. A full-featured but difficult to use TEXT interface allowed librarians and a few medical researchers to access MEDLINE. A graphical interface, with fewer features, called Grateful Med allowed doc-

tors, students and researchers to search MEDLINE themselves and to request via the software that a library send them an article they wanted for a price. The marketplace licensed the MEDLINE data and several firms offer alternative software that exploits the robust indexing and notable full strength of the MeSH headings. These offerings were good enough to compete successfully with NLM's interfaces. Enter the click.

Neither of these interface strategies however could compete with the hunger brought on by the click for direct access from citation to item cited. Neither could compete with the immediacy of AltaVista and yahoo. Both NLM and MEDLINE services needed direct access to the journal articles. But the publishers were concerned. Would their market hold up to digital access? So, one vendor, now called Ovid Technologies, wrote agreements with publishers to get them to open up access: Ovid would re-key the journals, would not offer access before the print versions were available, would place in restrictions on use, and would pay the publisher well for the content. Ovid offered the first MEDLINE with clicks to the articles; but others followed soon after. Another vendor, OCLC's FirstSearch has its version of MEDLINE. But it has deals with a different group of publishers. So when you make a search in each – you get different full text. In fact, a search in Ovid may show no full text while one in OCLC may have a dozen full text articles.

And publishers began to see the possibilities in a less threatening economic light. As the publishers themselves opened up web sites – by subscription or on a trial basis – NLM saw the possibility to use its content to provide navigation across all medical publishers. Today MEDLINE offers links to about 400 journals from the citations in the database through its PubMed program. Furthermore, MEDLINE actually has the publishers providing the skeleton citations for the MEDLINE database, speeding up availability of citations, and article clicks.

Enter the click. At the American Pediatrics Association the possibility of providing its journal through Ovid seemed like a good, safe way to let users click to it. But as other publishers and associations began to put out their journals on the web, the APA decided that it too must move with the time. And like other agencies, just webifying the journal was not enough. And so the

editors hit upon a wonderful scheme – they would take those articles that were too long or not quite as good (and therefore did not make it into the print version) and publish these on the web for the APA members. But how would the members know about them? They would publish an abstract and a URL in the paper journal to guide readers to the site. The first issue based upon this model debuted in December 1998. Now the Ovid version of *Pediatrics* was not only slower to come out, but also did not include these articles and citations. But by participating in the PubMed program, *Pediatrics* sends the citations to these articles along with the paper-printed articles' citations to MEDLINE so that they would be indexed.

Enter the click. The library in an effort to provide clickable MeSH enabled access to medical journals had licensed several hundred titles from Ovid. Among these was *Pediatrics*. On December 5, 1998, the first user located a citation in MEDLINE for one of these web format, APA only, articles and was surprised that there was no access from within Ovid. Fortunately, the librarian did not immediately assume that this was because of Ovid's slowness to post journals due to the embargo on not making their electronic version available ahead of the paper version. Instead she located the information about the new system at the *Pediatrics* web site and determined that neither the library's print subscription nor its electronic subscription would provide access to the requested article. Only a subscription directly with *Pediatrics* could be used – the licensing of content for groups of people versus managing the password access for individual subscribers for an institution and other similar matters could form the basis for another talk. But the lesson we are concerned with is that despite their appearance not all clicks – even to the same authorised content – are equal.

### *Licensing information*

This inequality of content – even similar content – has raised some interesting issues for libraries. As libraries try to provide access to the growing number of high-quality information resources available by subscription license over the Web, we are struggling with major policy issues on behalf of our users.

With each license, we consider carefully the access rights. Who is the legitimate user community? Does the information provider allow student, faculty and university staff to use the information from their homes? Can their families use the information? Can all the people who are allowed to use the library have access on the library's site? On the campus? Are there limits on the uses of the data? What about long term access to the data – does the provider have a plan to archive the data and to make it available? Can the data be use over a network? Are there other limits to the uses – can it be re-posted as a classroom reading? Can the library use it to fill interlibrary loan requests? What happens if the user community infringes on the licensing agreement – who is responsible for infringement? Will the entire community lose access? What kind of usage data is kept? What about the privacy and confidentiality of the users of the information? As for the content: what is its relation to any print versions? How current and up-to-date is the online information? Does the online information appear before, after or at the same time as the print information. As for the interface: is it easy to use? Does it add value to the information? Are the functions complete and do they all work? What is the relationship to other resources – can you link to other items? And finally there are a whole bunch of miscellaneous items: the stability of the resource – is it available? Is response time good?

And as the librarian asks these questions, it isn't solely with the library in mind. These issues are important for the user. The library will warn the user of the weaknesses and promote the strengths. The librarian is fighting to protect the privacy of the user and their right to anonymous access to information.

### *The click*

And so we return to the click. This simple mechanism has transformed us all. It carries with it many assumptions, about access and economy to name two. These assumptions do not reflect the information research ethic that we in the university have embraced and taught for the last half of the 20<sup>th</sup> century – the reliability of citations and availability of information is no longer a yes or no answerable question, instead it is a maybe or

if. The click – now slightly more than 5 years old – has undone all.

This takes us to the critical question of what are the underlying features of the click? How will they affect us in the future?

Let's look quickly at four features of today's click: navigation, identification, commerce and enforcement.

### *Navigation*

Navigation is the most clearly understood – the way to get from here to there. It mostly works except for when it doesn't – about 40% of the time. Servers are up and down. Routes are open and closed. Traffic cops are working or not. The average life of a click on the web is about 60 days. But the protocols underlying the click seem robust and, for most of us in universities, we can easily and quickly retrieve textual matter.

Navigation through clicking is so powerful a metaphor that users now expect to click on any underlined text in any document. They are frustrated and annoyed when there is no hyperlink embedded. It's as if underlining for rhetorical emphasis has lost its purpose.

The fact that the navigation may be to some place we don't want to go seems somehow unimportant. It's that joyful leap from one place to another that symbolises the freedom to explore. Whatever its attraction, if you can use a mouse, you can click, whether you are knowledgeable enough to read the text around the link. Whether that click will be there next week or even tomorrow – you can't know.

How we find and choose our navigation clicks is an important issue. The ease of use and comfort with the technology mean that whether the World Wide Web is the best place to get an answer or not, it is the place that we will turn. In the early days of graphical business databases, undergraduate students could be seen using them to research papers for literature courses. This wasn't because they found good sources, but because the clicks were a lot easier than leafing through paper volumes of the *MLA Bibliography*. One wonders what the students did with all those articles about investing in literary works and about the prices that authors' autographs fetched in auctions!

### *Identification*

Identification is a bit more problematic. How we tell that the item we saw last week is the same one we see this week – the same address is not necessarily the same content and identifying objects is an important issue being discussed worldwide. Clearly this is complicated when in fact there may be three definitive editions of each issue of *Pediatrics*: *pediatrics.org*, paper and Ovid. Citation is an important part of the scholarly process – how will we respond to this issue?

Must we print out everything we cite? Must we download everything we use for our scholarship? How can we be sure that we will be able to reassemble those bytes that seemed so compelling? The research dealing with universal resources names, digital object identifiers, SFX technology, standard numbers for electronic resources, purls, handles, and other schemes is promising, but only in its infancy.

Furthermore, what are the clues that help us identify good information? Where are the worn paths that take us to the important address on the Web? The Web's full embrace of constant change means that even old friend sites may be unrecognisable after technology facelifts. How can we tell if this is the site used by a colleague?

The identification of resources, the referral of colleagues and students, the reliance on the content – all familiar in the print world – have been so transfigured in the electronic world that it should leave the researcher breathless.

But navigation is so easy, and, sometimes, so fast, that we mistake the familiar click for identification.

### *Commerce*

Commerce seems to be as important as the technology. And clicks are now smart enough to carry with them the information that you may have access (because of your Internet address, or because of information stored on your hard-drive in a cookie). It nice that sites you use often recognise you and can alert you to things of interest – a cheap airfare, a new book, a contest, a sale. The Web suddenly begins to change as you see more advertisements that seem to be related to your interests – computers, gardening, and travel. Is it that the Web is a place for people with your likes,

or is it that the Web knows what you like and is trying to change to be the place you want it to be.

This move toward commerce-ready clicks also may carry some information that could link use of an item to you personally. It moves one away from the anonymity of the book stacks or the library's linked public computer. Instead the user who searched for a "hotel in Monaco" can discover minutes later that ads about the Monaco make of automobile are appearing on her screen. Even six months later, an email will arrive suggesting that you might be interested in some special travel programmes to visit Monte Carlo.

The aggregation of your interests, and the passing on of your profile can create both an improved, more relevant Web, and also a more narrowly defined one – limiting your choices.

### *Enforcement*

Finally, as content is licensed to you personally – forever or for an hour or a few days – the click contains the information to enforce your access or deny your access. The click can put information into the copy you download or print so it can be traced to you – even second generation copies may bare the marks of the source of the item. Ownership can be a very good thing, especially to identify that something belongs to you. Think about the bookplates you may put in your books in the hopes that friends will return them. They identify you with this item. But it's one thing to choose what you display in your home or admit to owning when you lend it to a friend, it's another thing when that item is being tracked and your personal use being reported. In a print world, once you've bought the book the physical object is yours. You can pretty much do as you please with that one item without notify-

ing the publisher or author as long as you don't republish it. In the world of Web commerce enforced clicks, the action of buying the item becomes a piece of information itself that can be sold, traded, or used in other ways. The item can report your use of activities like printing – because this activity can be both legal (one copy for you) and illegal (many copies for others).

The latter two issues – commerce and enforcement – are particularly troubling to librarians. Our systems have been built to separate the navigation and identification from the commerce and enforcement so that individual users information habits can not be tracked. But the click has been seducing our users into a false sense of euphoria over quick access that is neither complete nor secure.

### *Conclusion*

For librarians the ethics of the click is a serious business – and I hope you can now see why. If you can see why, then next time, in the split second before you click, you'll ask yourself, am I clicking because it's there or is there a better place to get my information? Maybe you won't stop, but if you just pause, the revolution against the mindless click and the formation of a community of educated clickers has just begun.

### *References*

- National Library of Medicine. 2000. Database Update Information URL: [http://igm.nlm.nih.gov/splash/IGM\\_dbupdate.html](http://igm.nlm.nih.gov/splash/IGM_dbupdate.html)
- Romano, Frank. 1998. "Will printing disappear?," *Electronic Publishing* 22(10): 26.
- Zakon Robert H. 2000. Hobbes' Internet Timeline v5.0. URL: <http://www.isoc.org/guest/zakon/Internet/History/HIT.html#2000s>