

The (re)creation of scholarly journals: Document and information architecture in Open Access e-journals

Short paper for DOCAM '06,
Berkeley, 13-15 October 2006

Helena Francke
Swedish School of Library and Information Science
Göteborg University and University College of Borås
SE-501 90 Borås, Sweden
Tel. +46 33 435 43 20
Fax. +46 33 435 40 05
helena.francke@hb.se
<<http://www.adm.hb.se/~hfr/>>

Introduction

“How is the scholarly journal remade in electronic form by a partly new group of producers?”

With this question as a point of departure, I will focus my talk today on how an approach to document analysis that will focus on design and media-specific properties of electronic journals can be designed. The presentation is a description of and reports some preliminary results from my ongoing PhD project.

There has been a hype around scholarly journals in electronic form since the 1960's [1] but it was with the introduction of the World Wide Web in the early 1990's that electronic journals really became an important means for formal scholarly communication directed at end users. To be able to access and download scholarly articles directly from one's office, one's home, or a hotel room on the other side of the world is of course extremely convenient. The access may come with restrictions, such as a requirement that one identifies oneself as a registered user at one's university library, or pay a fee to access the article, but in many cases, the journal articles are made available free of charge. Those are the journals that I will focus on today.

Despite the convenience of web-based journals, the use of a new medium for scholarly communication is not unproblematic. When I discuss e-only scholarly journals with researchers and students, particularly journals that are not produced by publishers that are well-known to these people, they often express similar concerns:

- Can I trust online journals?
- Can I trust it not to disappear?
- Is it indexed in my library's OPAC?
- Isn't there a risk someone will just copy my text and use it?

As you can see, these questions are primarily questions of trust. And trust in this case has to do mainly with the material and technical aspects of the journals, with the choice of distribution and storage technology. Over the years, the print scholarly journal has established itself as a fairly standardised – and in that way predictable – source of scholarly knowledge. Now that journals are increasingly published on the web, they need to negotiate the properties that specifically characterise the web medium. The scholarly journal as a document type will need to be remade to fit the new medium if the e-only journal is to become a competitive alternative to print in the future.

It is quite likely that the intellectual content of the scholarly article, the way in which a scholarly argument is structured and constructed, will remain more or less the same in the immediate future, even though the multimodality of the web may encourage the development of new rhetorical structures and increased possibilities to refer to and include scholarly evidence. But our documentary practices [2], the ways in which we use and interact with the documents, change when we change medium. This is what is reflected in the questions posed earlier, where some properties of the print medium are perceived as lacking in the electronic medium: the flexibility of electronic media makes people more on their guard than is the case with print documents; the potential ephemerality of material on the web calls for new, secure solutions for long-term preservation; and whereas the journals are no longer physical objects of paper stored in the library, there is still need for them to be accessible and, not least, possible to locate and retrieve through institutions that people associate with credibility, such as libraries.

Remediation

The way in which media use and mimic each other's properties has been theorised by Jay Bolter and Richard Grusin [3]. Their view is that producers of a medium borrow from other media in a competitive struggle to present the medium in question as the most attractive alternative. The ways in which the borrowing – the remediation – is conducted could be said to be positioned along a spectrum, from a respectful way, where the medium is presented as resembling other media, to a more radical approach where the medium is being presented as something completely new in that it makes use of elements of other media in manners that offer completely new functionalities and experiences. This view of media emphasises that although there may be limits to what can be

achieved with a particular medium, there is not *one* given way in which it can be structured for storage and use. We carry over social conventions and practice from one medium to another. The web medium, for instance, could be used so that it closely resembles and takes over functions of other media: print, radio, telephone, TV or film. But it can also be presented as something new and different by emphasising the hypermediacy that comes from mixing elements from other media into interactive web pages incorporating text, sound, moving images, etc. The commonly asked questions mentioned before indicate that when it comes to scholarly journals, many people feel that the web medium lacks qualities found in print. On the other hand, other genres, such as the scholarly edition, have been criticised for not implementing the radical, hypermedial functions that the web medium offers [4]. How does the document type of scholarly journals remediate other media when it moves online? How does it make use of the web medium's material properties to reuse or renegotiate elements that over the centuries have become standardised in print journals? And could the online journal be said to contribute something entirely original to scholarly communication? These are the type of questions that I am interested in investigating.

Finding the journal population

People's concerns about scholarly e-journals come to the fore in particular with scholar-initiated e-journals that are available on an Open Access, or at least fee-free, contract. These scholarly journals have been described as editor-managed rather than publisher-managed [5] because all stages of their production are in the hands of the editors rather than some of them being managed by professional publishers. The fact that these journals are not benefiting from being associated with publishing houses that are considered to be cognitive authorities [6] makes them particularly suitable to study from the perspective of the questions mentioned earlier in this presentation. Is it possible that those journals that are only published electronically have a freer connection to the print journal form than publisher-managed journals do? Or do they try to reassure their readers by mimicking the print journal as far as possible? The relationship is complex, and for that reason, I have found the editor-managed Open Access e-journals an interesting group of journals to work with. They differ from the print journals not only with regard to storage and presentation media but, most often, with regard to who is in charge of the journal's material form.

The task of identifying and narrowing down the relevant population of e-journals is far from simple. To identify potentially relevant journals, I have used a number of existing databases that list Open Access journals or have Open Access as a limitation criterion. However, the additional requirement that the journal should be editor-managed made further limitations necessary. And what journals are in fact editor-managed? Individual scholars or groups of scholars seem to fit the criteria, as – mainly – do universities or university departments. But what about a

university press? Some of these are obviously established publishing houses, such as Oxford UP or MIT UP. But some publishers that are named “university press” seem to be run by individual scholars. And where do librarians fit in? Several libraries are engaged in e-journal or e-press initiatives. On the other hand, there are large-scale e-journal projects such as British BioMed Central, Latin American SciELO, or Japanese J-STAGE, which are either to be regarded as regular publishers (BMC) or which provide similar e-journal interfaces to a group of mainly print publications (SciELO, J-STAGE). I have chosen to be fairly inclusive in my selection, and to also include e-journals that have print counterparts, but to exclude the large-scale projects, since their e-journals can be regarded as publisher-managed rather than editor-managed.

Document analysis investigating document architectures and information architectures

With focus on the editor-managed Open Access journals, I want to investigate expressions of practices, that is how the documentary practices of authors and editors are manifested in the journals with respect to material document properties such as the journal’s design and use of media-specific properties. My focus will thus be on what can be found by studying the documents rather than studying the people working with the documents. For this approach, I have chosen a framework for document analysis which is designed to investigate document architectures and information architectures. In this context, *document architecture* concerns how a document’s constituting elements form structures that determine the internal organisation of the document. These structures could be formed for instance by the files that create the document, by the internal positioning of rhetorical content, or by how the document elements’ are placed in the layout. *Information architecture* describes how the document’s intellectual content has been categorised and labelled and how it is made physically accessible to the user by navigation of the document. Such an architectural framework entails a view of the electronic document as an artefact with inscriptions of intellectual content that are organised in complex structures at different levels. More specifically, the architectural framework allows me to highlight such properties of the electronic document as file format, mark up and visual design, metadata and other types of content included and its order and positioning on the site, as well as how the site can be navigated and searched.

The framework needs to be modelled in some way so as to form the basis of an empirical study. I have chosen to restrict myself to manual rather than automatic collection of data, but to include methods that are both qualitative and quantitative. In order to indicate tendencies and trends in e-journal publishing, I am conducting a survey of e-journal architectures. The qualitative study of a smaller number of journals is intended to provide a more in-depth understanding of the journal architectures than can be reached through a survey. It will take the

form of ‘close readings’ of the journal web sites, readings that focus on architectural components rather than on intellectual content.

My initial question suggests an element of comparison in my study. So far, I have spoken primarily of online journals, but they need to be placed in relation to the architectures of print journals, and perhaps also to publisher-managed e-journals. With regard to print journals, I intend to gather data partly through a literature review of previous work and partly by exposing a few print journals to a qualitative study similar to the one just described.

Performing document analyses

I would like to go on to briefly outline how findings from a qualitative study and from the survey can support each other. As examples, I have selected the journals’ choice of file formats and their use of different modes of representation.

The figures from the survey are as yet preliminary. They only represent a sub-sample of the intended final sample, and can at the moment represent only the journals that were part of the study, not be generalised to the entire population. Nevertheless, I think that they point to current trends. The sub-sample contains 155 journals that were selected as a random sample from the previously identified population of approximately 700 journals.

I will also use examples from two journals to illustrate the qualitative part of the study. These journals will probably not be the ones used in my thesis. In fact, for this presentation I have chosen one of the examples – *Human IT* [7] – because I have worked with it and therefore know it fairly well from the “inside”. This of course makes it different from the journals I will work with in the thesis. The other journal, *seminar.net* [8], is selected because it is one of the few journals that consistently use video as an added value function to their articles.

Example 1: choice of file formats

Human IT, a journal for studies of ICT from a human sciences perspective, started in 1997 and was from the start published as a parallel publication in print and on the web. Web distribution seemed to suit the computer-savvy, curious, and open-minded audience that was envisioned. Yet, when a few years later we began to talk about discontinuing the print version of the journal it met with resistance from many people. Their concerns were similar to those I mentioned earlier. To these concerns were added the issue of reputation and credit. The print journal was clearly associated with academic status to a higher degree than the e-journal.

Up to the point when the print version was discontinued in 2002, the e-version of *Human IT* had been published as HTML. When the print version was abandoned, the web version of the journal was adapted to resemble the previous print version even more closely. PDF substituted HTML as the publishing format for the articles and the page design was made to mirror

that of the print journal. The main reasons for this choice was that the length of the average article in the journal made it likely that many readers would prefer to print out the article before reading it, and that in these cases PDF would be better at retaining a reader-friendly layout. To provide the readers with both formats was considered to make the publication process too time-consuming, given the available technology.

As it turns out, *Human IT* seems to be fairly typical of many editor-managed open access e-journals in this respect. Presumably, the fact that PDF has now incorporated such HTML features as hypertext linking and that the format is retrievable through search engines has lessened the attraction of HTML somewhat. The survey, of course, shows only the current file formats available; they could have been changed over the years. I have looked at the first article in the earliest issue and the first article in the most current issue included in the web version of each journal. There is an indication of an increase in the number of journals that publish their articles in PDF. Looking at the format used in the first issue, by year when the issue was published, seems to indicate that HTML was more popular before the turn of the millennium, but that it has then lost its dominance to PDF.

This is supported by looking at the articles in the journals' first online issues: 42% are published only in PDF. In comparison, 56% of the journals use only PDF in their most recent issue. For HTML, the ratio drops from 34% to 23%. If one adds to this the c. 21% of journals that publish their articles in more than one format, I found that 59% of the journals publish articles in PDF in the first issue and 74% in their latest issue. HTML usage falls from 49% to 39%. This indicates that an increasing number of e-journals publish their articles in PDF, but also that about one fifth of the journals provide more than one format, usually both HTML and PDF. Those two file formats are completely dominating: none of the journals had published the first article in their most recent issue only in a format other than PDF or HTML. However, there were some examples of journals that also use for instance PostScript, MS Word/RTF and various (La)TeX output formats.

Example 2: choice of modes of representation

That the e-journal remediates print is not surprising. But to which extent does it remediate other media? This has to do with the modes of representation that are included in the journal. I make a distinction here between *media* and *mode of representation*. It is intended to capture the difference between on the one hand the mode by which the inscriptions of a document are stored and presented, such as paper journal, TV, and computer (cf. [9]) and on the other hand the semiotic modes of representation (cf. [10]) that take the forms of for instance alphanumeric text, photo, or video. Most of the 155 journals that I have looked at include some form of figures. Only 16% have

only alphanumerical text and no figures in any of the original articles in the most current issue, whereas approximately half of the journals have more than two types of figures in the first article that includes this element. Most of the figures are in modes that are also often found in print articles, such as tables, illustrations, diagrams, and photographs. Far fewer of the journal articles contain modes of representation that cannot be included in print but that are common in other media, such as video (2 journals) or audio (3 journals).

One of the journals that make use of video is *seminar.net*. The journal was first published in 2005 and is published only in electronic form on the web. It aims to “raise critical issues as to how ICT and media is used in education and in society, in order to promote democratic values, equity and justice in the field of education” ([11]). In each article, *seminar.net* include a video clip, usually about 3-5 minutes long, of the article’s author speaking about the article in the form of an extended abstract or an introduction and background to what they have written. In one case, a slightly longer conference presentation is included. The presentations in the three issues so far (the ones that I could access) are exclusively portraying the author speaking. Usually we see the author’s head and shoulders as he or she speaks into the camera. According to the journal’s editor, the videos have met with appreciation among the audience, but the journal’s attempts to get authors to include video sequences as part of or as illustrations of an argument in an article have not yet been successful [12]. Obviously, authors are as much key actors as are editors when it comes to bringing about a change in scholarly publication by embracing new modes of representation, modes that remediate other media than print.

Signs of remediation

This is not the occasion to draw definite conclusions from these examples. They indicate that the online scholarly journal is actively remediating many of the properties of print journals, while being more cautious in remediating other media. Remediation of other media is most likely taking place, and I am sure that you can provide me with many examples of this, but the main arena is not the scholarly journal. Rather, these two examples indicate that the online journal remediates the print journal in a respectful manner. With a more extensive and detailed study, I want to show how our conventions and practices are carried over from documents in one medium to documents in another, but also to offer a nuanced understanding of how these conventions and practices change. I am curious to see whether further investigations will provide a more complex view of the ways in which aspects of print journals are carried over to web journals.

Notes

- [1] Tenopir, Carol & Donald W. King (2000). *Towards Electronic Journals: Realities for Scientists, Librarians, and Publishers*. Washington, DC: Special Libraries Association. pp. 22f.
- [2] Frohmann, Bernd (2004). *Deflating Information: From Science Studies to Documentation*. Toronto: University of Toronto Press.
- [3] Bolter, Jay David & Richard Grusin (2000). *Remediation: Understanding New Media*. Paperback ed. Cambridge, MA: MIT Press.
- [4] Karlsson, Lina & Linda Malm (2004). "Revolution or Remediation? A Study of Electronic Scholarly Editions on the Web." *Human IT* 7.1: 1-46. <<http://www.hb.se/bhs/ith/1-7/lklm.pdf>> [2006-10-03]
- [5] Kaufman-Wills Group (2005). *The Facts about Open Access: A Study of the Financial and Non-financial Effects of Alternative Business Models for Scholarly Journals*. Clapham: The Association of Learned and Professional Society Publishers. p. 5. Available from <<http://www.alpsp.org/publications/pub11.htm>> [2006-02-25]
- [6] Wilson, Patrick (1983). *Second-hand Knowledge: An Inquiry into Cognitive Authority*. (Contributions in Librarianship and Information Science; 44). Westport, CT & London: Greenwood Press.
- [7] *Human IT: Tidskrift för studier av IT ur ett humanvetenskapligt perspektiv*. University College of Borås. <<http://www.hb.se/bhs/humanit/>>
- [8] *seminar.net: Media, Technology & Lifelong Learning*. Lillehammer University College. <<http://seminar.net>>
- [9] Gunder, Anna (2001). "Forming the Text, Performing the Work: Aspects of Media, Navigation, and Linking." *Human IT* 5.2-3: 81-206. <<http://www.hb.se/bhs/ith/23-01/ag.htm>> [2006-10-03]
- [10] Kress, Gunther (1998). "Visual and Verbal Modes of Representation in Electronically Mediated Communication: The Potentials of New Forms of Text." *Page to Screen: Taking Literacy into the Electronic Era*. Ed. Ilona Snyder. London & New York: Routledge. 53-79.
- [11] "Scope." (2005). *seminar.net*. Lillehammer University College. <<http://www.seminar.net/about-seminar/scope>> [2006-10-04]
- [12] Nordvelle, Yngve Troye. E-mail to the author dated 2006-10-03.