

"Stratum continuum of Information: Scholarly communications and the role of university libraries"

Leo Waaijers
Delft University of Technology Library

Leo Waaijers, "Stratum continuum of Information: Scholarly communications and the role of university libraries." *Proceedings of the IATUL Conferences*. Paper 28.

<http://docs.lib.purdue.edu/iatul/2001/papers/28>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

Stratum Continuum of Information

Scholarly communication and the role of university libraries.

Keywords: scholarly communication, libraries, university presses, digitization, publishers.

Abstract

As a consequence of digitization universities have to investigate their scholarly communication process. In fact, this is a quest for values that goes beyond the issues of the day. Once found, these values operate as criteria for assessing competencies, roles and instruments.

The communication process must allow for free exchange of ideas and results. In particular, knowledge created in the public domain must be open accessible. The process must be reliable, lend credibility to authors and give support to users. The respective competencies are logistics, validation and mediation. New technologies do not only support these classical aspirations, but develop their own dynamics when it comes to broadening the communication scope. Publications can be enhanced by including audio-visuals, executable files and datasets. Linking techniques enable the creation of organically growing bodies of knowledge.

Who is the fittest for this scholarly communication job? Is it universities and their libraries and university presses, or could it best remain in the hands of global publishing monopolies?

A spirit of rebellion

“Scientific progress thrives on the symbiosis of creation and communication”, is not a catchy phrase to open a paper with. It is more a platitude, a ‘so what?’ statement. Yet, the rest of this article will deal with the consequences of this platitude for universities and their libraries.

To begin with, scientific creation alone is not enough for a university. It would result in what is called tacit knowledge. It is the core theme of current knowledge management theories that such tacit knowledge must be made explicit, hence communicable, testable and operational. Companies, especially knowledge intensive companies, are aware of that and see communication of knowledge as their most critical success factor. Creation and communication are the yin and yang of progress in science. Both constitute the rationale of a university. When society pays taxes to subsidize their knowledge generating institutes they may expect that the results are widely and freely available for further development, education and human welfare.

If knowledge communication is that important, it deserves a more systematic approach than universities were inclined to give it in the past. Universities have always organised conferences, given lectures, judged doctoral theses, provided editorial boards and referees, issued reports. They had grand libraries and petit university presses. These were all elements in a highly fragmented scholarly communication process and it was the contradictions of digitization that commanded a rethink of this historically evolved situation. While we, the universities and their libraries that is, expected that digitization would make scholarly communication cheaper, we were confronted with relentless annual price increases of over 10 percent. When we wanted to use the Internet to disseminate knowledge, we discovered that publications were imprisoned by monopolies. When we stimulated prior publishing of articles on pre-print servers, we learned that this could inhibit formal publication in a prestigious journal. In developing digital course ware we found out that we needed permission for using our own material. When we used scanned images to forward documents instead of sending hard copies by snail mail, we were warned that this would only be tolerated for the time being. Instead of moving into a free-er world, we entered a world of intricate licensing contracts bristling with legal pitfalls and new notions like authorized user, permitted user, registered user, walk-in-user, selected user, perpetual access, IP-domain and site, reasonable effort and best effort. All these terms without jurisprudence have one thing in common: the risk of breach or not meeting the requirements is always laid on our side. Warranties are never given by the supplier. Then, by way of unbelievable generosity, annual price increases were reduced to 5 percent. But only in exchange for a five years non-cancellations contract. And, a propos, the annual advance payment remained unchanged. All this occurred concurrent with a dramatic lowering of production costs on the publishers’ side caused by the new digital technology. The costs are even further reduced by enforced bulk purchase via mandatory mega licenses that come in place of the annual finetuning of individual library collections. Under even the

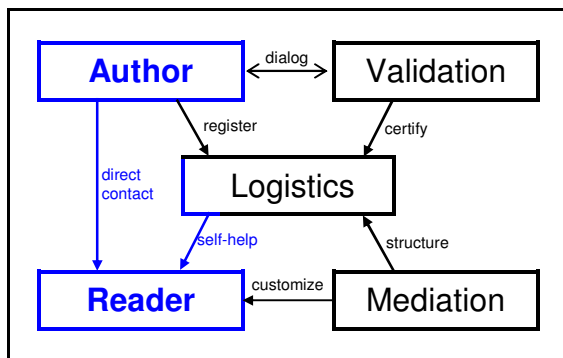
weakest market conditions this would be completely unthinkable. An extra side effect of the monopolistic control that publishers have is that the usage profile per IP-adress, that is per machine, or per password, which means per user, is scrupulously monitored. In the corporate arena of knowledge based competition, this is invaluable strategic information. Companies hardly seem to be aware of that.

It has been this experience of dependency, economically and legally, that has raised a spirit of rebellion, especially in American universities and their libraries, gradually followed by their European counterparts.

Core competences

A rebellion is best undertaken when it has an intelligent basis. Here, intelligence is found in the cultural values that underly the process of scholarly communication. On pain of failure, any alternative for the current situation should respect these values. What are they?

First and foremost, freedom of exchange of ideas and results. The more free communication is hampered, the more inefficient the process of knowledge generation becomes. On a wall in the Delft library a slogan says: "Information is the only resource that grows with use." Universities have known that for centuries already and modern knowledge management theories depart from there. Recently, MIT announced their Open Courseware project and decided to make public all their course material on the Internet. When I disussed this with Delft people, I only met twinkling eyes. Our scholars love the idea; in fact, they seemed to be somewhat jealous.



Reference model for core competencies in scholarly communication

Next, the communication process needs a professional infrastructure that validates authors i.e. lends them credibility, is based on reliable logistics and mediates for users i.e. gives them support. These three core competencies represent the classical values of respectively quality, secure storage and adequacy of information. They fuel the communication process, without them it comes to a halt. The values per se are culturally embedded and medium independent, although their appearances may differ as a consequence of the technologies applied.

Lending credibility cannot be done by any machine yet. It is an inter-human process within the scholarly community where, like in guilds, peers judge the results of apprentices. This process is as old as science itself, irrelevant of the information carrier. The main stream of the validation process is organised inside academic institutions like universities, research organisations and learned societies. Here, reports are assessed, doctorates conferred, books edited, papers admitted to a conference, tests marked, prizes awarded, articles accepted in a journal, documents included in a collection, references selected for a database. The corresponding professional activity is to organise and maintain open selection processes according to objective and methodologically stable accepted standards. Unfortunately, a relatively small portion of the selection process fell in the hands of a couple of global high-profit driven publishing monopolies. It concerns a maximal 2000 journals and in the ubiquitous university culture of constant validation and quality assessment they form an anomaly. What's more, the core activities of editing and refereeing these journals are in fact executed by the academic staff themselves and for free. The greed and globalization of the involved publishers is a relatively young phenomenon that had Robert Maxwell of Pergamon Press as its founding father and Pierre Vinken of Elsevier as his faustian pupil.

Next, information logistics is reliable when it registers accurately, secures authenticity, guarantees access and retrievability and safeguards availability. In the paper world accurate registration of a new publication could be improved. After submitting an article it disappeared for some time in a refereeing circuit unknown by the author. Especially when the article was rejected in the first instance, the accuracy of the claim data left room for desire. Securing authenticity was best met by mass printing, wide dissemination and distributed storage in library stacks. Cataloging took care of retrievability and

document supply through lending or photocopying safeguarded availability. In the digital world the respective requirements are met by timestamping and 'freezing' of the article, limited storage in networked repositories, automated indexing and attaching meta data, followed by open access, downloading and printing. Compared to the paper world retrievability and transport are superior in the digital world. But securing authenticity and ensuring durable storage are still inferior and in their digital infancy. Here paper has a long tradition.

Authors and readers do a lot of the logistics themselves by organizing seminars, making telephone calls, e-mailing etc. Professional parties in the world of information logistics are multitude, including for example post and telephone companies, internet providers, data warehouses etc. Confining ourselves to information logistics in a direct sense we find publishers, serials agents, database builders and libraries.

The third core competence, user support or mediation, is about structuring information according to profiles. This can be generic pre-structuring so that the user may help himself. But it is also specific ad-hoc structuring or customization in reply to a concrete request. Classical examples in the paper world are the open accessible collection, where the documents are presented according to discipline and within the discipline according to the alphabetic order of the authors. Other examples were card catalogs that offered access according to the first meaningful word in the title of a document or according to some keywords taken from a thesaurus. In the digital world all these things and more are integrated increasingly in virtual knowledge centres or portals. These portals do not only give complete documentary information but combine it with information about current research projects, institutes and fellow experts, availability of research grants and scholarships or study curriculum, fellow students, colleges, studentships, etc.

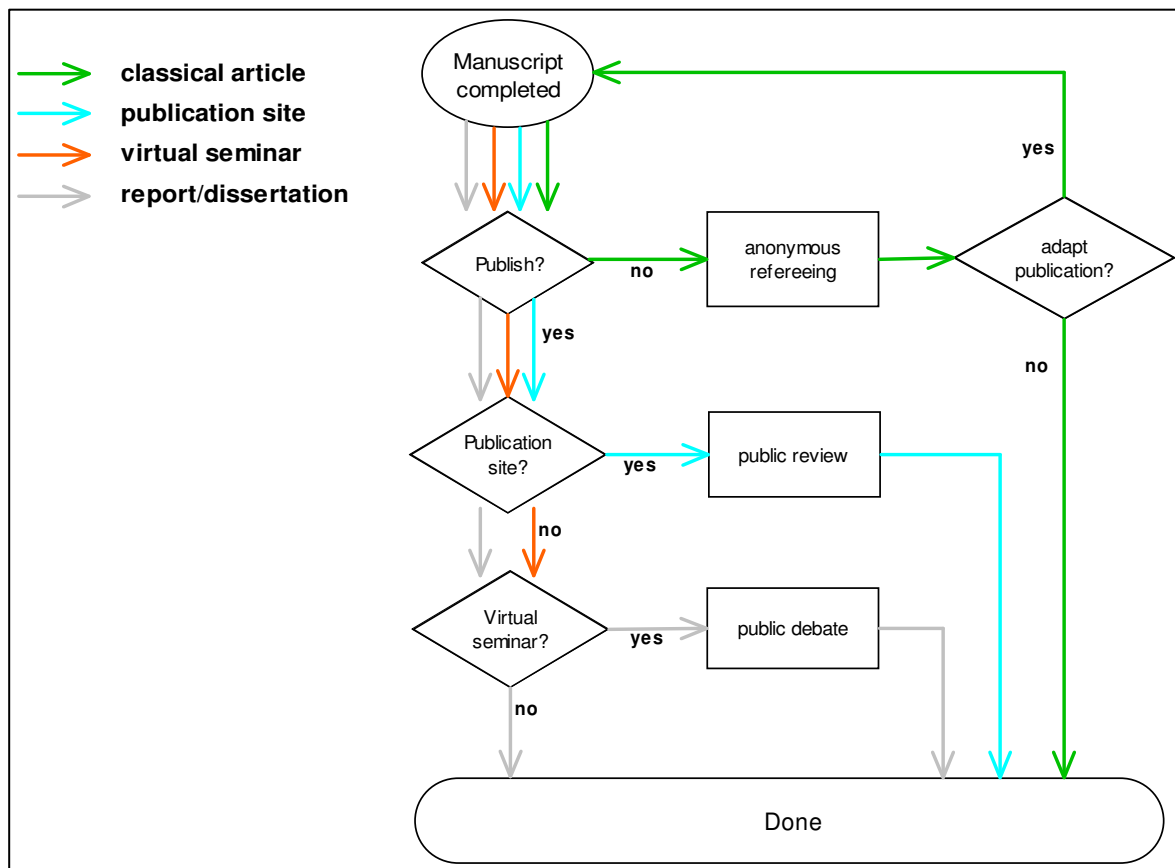
Mediation is successful if it catalyzes the user's creative process either because he is supplied with the right information or it evokes a serendipitous leap. As we have no clue of how the knowledge creating process works, the best practice here seems to be to offer what may be called information rich environments. The worker, be it student or scholar, then follows his own track, often targeted, sometimes associative. The structure offered should be as self evident as possible, but in the very dynamic information world we may help users help themselves by offering courses and training in retrieving and using information and writing web enabled documents.

Although academic staff or research units sometimes maintain their own paper or virtual collections, here libraries play the dominant role. They had a local monopoly, were not profit driven and were sometimes rather inefficient.

New developments

The above mentioned step of MIT is just an example from a broad range. The Sparc initiative is another example and the resounding success of JHEP yet another. The most recent signal is the manifesto for the Public Library of Science that took the the medical and life sciences scholarly community by storm. In a few months almost 25.000 scholars and students signed up for the establishment of an online public library that would provide the full contents of the published records of research and scholarly discourse in a freely accessible, fully researchable, interlinked form six months after their publication.

An investigation of the state of the art in scholarly communication in a more systematic way demonstrates that these initiatives to free intellectual academic produce go beyond disjointed incidents. The common denominator of the contemporary and broad variety of actions is the desire to take full advantage of modern information and communication technology to vastly increase the accessibility and utility of scientific literature, to enhance scientific productivity and to grant public availability and free distribution rights of the record of scientific research and ideas. Some fundamental developments that support this movement will be mentioned here.



Alternative scenarios for quality selection in the Dutch Roquade project

The main problem in the area of quality selection is its inefficiency. It may easily take more than a year for an article to be refereed and hence to be published. Modern technology enables alternative selection procedures whereby validation takes place after publication.

Two main approaches are open reviewing on a publication site and the virtual seminar. Open reviewing is commonplace for almost every intellectual product in our world. The piano recital is played first and then commented upon by professional reviewers the other day. The same holds for the new building, the stage play, the Queens speech or the State of the Union, the governmental report etc. The principal advantages are twofold: first, reviewers and reviews are known instead of anonymous and unseen; second, no copyright transfer is involved. A practical advantage is speed; publication is not suspended until the referee has given his approval.

In the virtual seminar an article is published on the Web and evokes a discussion between readers and the author, which may give rise to new versions of the original article. Here also, speed of publication and openness of the validation process are advantages over the classical process. And of course, no copyright transfer involved. An interesting aspect of the Roquade initiative in The Netherlands is that it supports both the classical refereeing process and the modern varieties of open reviewing and the virtual seminar, thus offering authors and editors the option of gradually changing from one paradigm to the other.

Looking into the logistical part of the scholarly communication process we encounter the Open Archives initiative (OAI), that carries on the legendary Los Alamos pre-print server. To day these archives can be combined with emerging open standards for smart tagging like open URL's, that constitute the basis for SFX type applications, and Dublin Core based meta data, feeding Web agents. The generic reference model for an Open Archival Information System (OAIS) constitutes the basis for an ISO standardization of digital deposit systems. Together these systems provide the sound and traceable infrastructure, the Stratum Continuum of Information, that form the foundation of our new validation and mediation activities.

Finally, new business models are under discussion. Under the current circumstances authors, or more precisely their research funds, are substantially granted on a formula basis, that is by counting the number of publications and citations. This is more an incentive for quantity than for quality and, among others, it leads to trial submissions of inferior articles and repeated recycling of the same information. As a consequence there is also a refereeing crisis, beside and less visible than the serials crises. There are simply not enough referees for maintaining a high standard refereeing proces, all the more so because the referees are also researchers who are compelled by the same system to publish more and more themselves. This gives the system a constant inflation rate. An effective feedback mechanism would be to simply charge authors, that is, their projects or research unit, for having their articles refereed and then reward the referee, again his project or research unit that is, for the job. It would make authors more conscientious and stimulate referees, thus creating a better balance. Be aware, this is not a payment for being published, like current page rates. It is a payment for being refereed and the fee is owed whether your article is accepted or rejected. It's like an examination fee and a relatively small portion of the research budget, somewhere between 0.5% and 1%, would do the job. As the fees flow back to the referees, for the research community as a whole the operation is budgettary neutral anyway.

Other steps of the economics of the information chain should be investigated as well. What we basically need is an activity based costing analysis of the whole scholarly communication proces, from producer/author to consumer/reader. It would provide us with a fundament for the debate about the alternative business models and decisions for future scenarios. However, this is beyond the scope of this paper. Would it be an idea for IATUL to commission such a study?

Body of knowledge

So far in this paper the word 'article' has been used without further reference, as a known entity that needs no further observation. However, looking a bit deeper into the notion reveals questions concerning the nature of the document. In the paper world an article is a printable and linear piece of information, with an introduction, a body and a conclusion. In the digital world, the article is freed of the limitations of the need to be printed and may contain audio-visual elements, executable programs and big datasets. They give the document new dimensions but leave it in tact as a unit. The same can still be said of the Web enabled interlinking of documents. References in a document to other articles, to authors, to subjects etc. can be made clickable, complete portals may be built, but again, these steps leave the target to which one refers in tact.

Now, XML is a new breakthrough. This radical new technique facilitates the anatomising of the internal structure of the document. Paragraphs, figures, formulas, pictures, quotations, conclusions etc. inside the article can be coded separately such that the code represents metadata about the content of the fragment. The concluding step is the interlinking of these fragments, thus using them as building blocks for new 'documents'. In its classical, primitive form one might think of quotations in an article, but these modern interlinking structures open a new world of non linear associative wandering. Thus, the student's cognitive process can be better supported than by classical linear reading.

In this context, several intriguing questions arise. Questions that are directly related to the three core competences validation, logistics and mediation.

Who should build these associative links? Are they the information specialists in a library, in which case we can dig up the good old reference librarian again? Or are the authors themselves responsible? But why should they do it; what is their incentive? Or is a new profession emerging, the 'information- or knowledge-associator'? And is the link builder a writer, a validator or a mediator? Here professions and competences seem to merge. In any case, building these links is an intellectual activity in its own right, that includes the quality question. The maker deserves recognition and links as such can be cited. Famous links may even be named. Joost Kircz, a Dutch information scientist, phantasizes about the one ultimate link that is awarded the Nobel Prize. On the other hand, the most simple links, i.e. those corresponding to litteral similarity or synonyms in text fragments, can be built automatically. Current indexing software seems a good starting point for this.

With this new XML technology it is not too difficult to foresee the construction of 'knowledge domains' that grow organically with increasingly intricate internal structures. The logistical question is: how to archive these intertwined complex knowledge structures that are composed of millions of information granules distributed over hundreds of servers? For this, we have to rethink the whole concept of archiving. Instead of safe-keeping separate entities like articles, we must think in terms of preserving organic information structures that are constantly developing, like an information biotope. As multiple

copies of essential pieces of information are stored, these structures should be able to recuperate when damaged. Software like Highwire's LOCKSS might be a good initial impetus in this direction. The notion of 'body of knowledge' will really assume its full meaning. From this point of view it is highly questionable if archival concepts that are based on perpetual preservation of digital originals in combination with iterative retro-emulation of original hard- and software, will be compatible with such an organic knowledge structure. To me, it seems that the emulation approach is only applicable for a relatively small number of individual cultural heritage documents: digital hieroglyphs in combination with the Stone of Rosette.

The third 'body of questions' concerns mediation. For healthy growth, a body of knowledge must be able to incorporate all relevant information. In other words, if a user, when wandering around in such an information rich structure, is repeatedly confronted with dead ends, the whole structure becomes useless to him, or even worse, an annoyance. As you know, currently, publishers form closed DOI-consortia, where they refer reciprocally to each others articles. DOI stands for Digital Object Identifier, a meaningless but unique number that refers to an individual article. It is the counterpart of the ISBN for books and the extension of the ISSN for journals. Here, 'closed' not only refers to the proprietary character of the DOI codification itself. There is more. It also means that a reader of an article, when clicking on a reference he encounters, will be faced with a stop if his institute does not have a license from the current owner of the document that is referred to. This might even happen when the referred document is written by a colleague from his own institute. A fortiori, it will happen when the link refers to an articles of his own hand! Vice versa, this holds for an author. When writing an article he will not be able to build in a link to another article unless his institute has licensed access to that document from the current copyright owner. This current owner could well be a high-profit publisher and the institute might be unable to afford the mega license.

All this is happening to day. Problems of this type will be intensified when it comes to applying XML techniques to build the aforementioned bodies of knowledge. Does one need the publisher's permission for anatomizing an article and incorporating it in a body of knowledge? Or is this just a more or less extensive use of the generic right to quote parts of any published work? If the latter is the case, it is not a risky prophecy that this interpretation will come under constant attack from the publishers and their affiliates, like copyright clearance centres, as is the case with current fair use clauses. The real threat is that the scholarly community will be divided into two parts: one of scientists that work in institutes that have licensed full access to increasingly intertwined information structures and those who may only wander around those structures and have an occasional clue of the treasures inside. In the world of scientific information the digital divide is a more serious threat than in society as a whole. The point is that in society government feels responsible for making publicly available the information that they produce. Their basis is both formal, because the generated information is paid for by tax money, and cultural, because this information is perceived as the blood stream of democracy. Although universities operate in the same culture, it is only recently that the academic community has become aware of the vital role of open knowledge dissemination for human welfare and of their own responsibility in that respect.

To conclude

To the libraries I would say, do not relent in making your institutes aware of their responsibility for the scholarly communication process. Meanwhile, "walk the talk", that is substantiate your own role in the complete information chain, especially with respect to the intellectual produce of your own university. Start archiving and publishing its dissertations, reports and conference proceedings, next organise quality selection processes for new journals in niche areas and then, make the big leap forward to the currently monopolized world of the hard core sciences. Of course, if there is a university press, amalgamate with it. Further, partner with colleagues and use open standards .

To the publishers I would say, forgo copyright transfer, start competing based on added values, which are quality selection, editing, efficiency, timeliness and presentation. Simply, realise the best price performance ratio instead of exploiting a monopoly. Then, offer your services to the academic world. This will be your only chance to come to terms with the scholarly community. But hurry! Otherwise they will have organised themselves. New technology and open standards are their opportunity; your conservatism is their incentive.

Leo Waaijers
University I Librarian Delft University of Technology.