ScieCom info
Svenskt Resurscentrum
för vetenskaplig kommunikation
www.sciecom.org

Open Access and Journal Publication in the Social Sciences and the Humanities

Ulla Carlsson, Dr, Director of Nordicom (Nordic Information Centre for Media and Communication Research) at Göteborg University
ulla.carlsson@nordicom.gu.se

Open Access opens up many new opportunities for communication among researchers and can help to make research and research findings visible and accessible, thereby enhancing the impact of the work in question. This applies not least at the national level, particularly with regard to doctoral dissertatations, monographs, conference papers, journal articles, etc. What is needed are readily accessible high-profile channels of good quality, and the universities and members of the research community alike have a major responsibility in bringing these about. Reaching one's target readers is decisive. A failure to establish that crucial link means failure for the entire endeavour; there can be no research communication, further accumulation of knowledge or broader collaboration without interested readers and users. Open Access can already point to many positive achievements — *DOAJ, Directory of Open Access Journals*, for example — that have opened doors and windows.

Open Access can also help to spread Swedish research in the international arena. But making research and research findings available is not enough. Again, the crucial link is scholars abroad who take note and make use of the work. There lies the proof of its merit. The quality of scholarship is validated when scholars

operating in other, foreign arenas value it as a positive contribution.

Visibility in the international research community

We confront a question there is no getting past: How can Swedish researchers in the Humanities and Social Sciences initiate dialogues on the international scene? How can a little country with a language spoken by only nine to ten million people gain recognition? Sources of research funding in Sweden today stress the importance of publishing in “reputable” international journals. As a medium, the journal is well-established, and it is difficult to envisage a replacement. Similarly, although the prevailing system of peer review may have its faults, we have yet to see a better alternative. It is important, however, that the research community strive to assert its interests and assume a greater measure of control over the journal market. That is to say, researchers must attain a greater measure of autonomy vis-à-vis journals than they have today. Secondly, they must work to ensure free access to articles via the web — through, for example, quality-assured open access titles. These two steps can render the market for scientific literature more accessible and augment the flow.

It is also important, however, to create new effective channels of communication that reach out into the world, to gain recognition as actors in the global scientific system. One way is to contribute to forceful and platform-like channels on the web through Open Access, channels that are mindful of maintaining high standards. Prerequisite to attaining this objective is participation in a specific discursive community of scholars. Membership in such a community affords familiarity with — and in time a command of — the conventions of various genres, such as the journal article. Both peer review and the mechanics of citation indices largely operate on the basis of shared discursive conventions.

Studies have found that Nordic research has attained quite some “visibility” in the *Social Science Citation Index*. Scandinavia and other small countries, e.g., The Netherlands, have been successful in “breaking” into the Anglo-American hegemony, whereas countries like Germany and France remain more or less excluded (Ingwersen 2000). Scandinavian scholars in the Social Sciences have assimilated the conventions of Anglo-American social science, for when we speak of international publication, we are talking about Anglo-American publication. The extent of “publication” differs quite widely, however, between disciplines.

Disciplines like Psychology and Political Economy have developed standards of scientific quality, which are reflected in the international flora of “quality journals”. The same applies to sectors of Political Science and Sociology, albeit relatively few Swedes in these fields are published in international journals. Estimations of

the standard of the work done in subjects like Social Work, Journalism, Gender Studies, Information Science, or Literary Studies vary widely — which is also reflected in the amount of work published in the respective fields.

Studies have found that the closer to the Humanities a given journal is, the greater the number of rejected articles. The chief factor behind this pattern is a relative lack of consensus regarding standards of scientific quality. A high proportion of rejections indicates a discrepancy between the editor's/reviewer's idea of scientific quality and that of the authors. Given that formulas for the journal article presume to reflect the conventions of the genre, it follows that texts that deviate from the model are lacking in one or another respect — e.g., badly written, poorly structured — in the eyes of one's "peers" (Swales 1986, White 2004).

In the Humanities, in many disciplines it is uncommon for researchers to take part in dialogues with members of other disciplines or colleagues in other countries. This is true of some disciplines in the Social Sciences, as well. In these areas, absence from the international scene may reflect absence from the national scholarly community, which, for that matter, is becoming more and more international in many respects. Swedish researchers have developed small discursive communities, which in turn have developed some parochial genre conventions that are not in alignment with the conventions of the international community.

Promoting research communication across frontiers

These are issues that must be discussed in every research funding institution and post-graduate programme in the Humanities and Social Sciences if we are to see any improvement in the standard of our research in terms of quality, relevance and the interest it arouses among international publishers. Publishing through Open Access can facilitate the process. In the longer term Open Access may in fact have a tremendous potential to augment international research communication to the betterment of research.

One main reason why Scandinavian researchers come out rather well in citation studies like the above-mentioned is the Norwegian publishing house, Universitetsforlaget, which puts out numerous widely respected journals in the Humanities and Social Sciences (Ingwersen cf. 2000). A concerted effort to create new, or to develop existing journals and other fora into similarly well-reputed channels on the more fruitful Nordic or European level may improve the current situation and build bridges to the international research community. Such an effort requires considerable work, commitment, creativity and intensive networking — but it will be amply rewarded.

I know this from my twenty-five years as publisher of the journal, Nordicom Review, an outlet for media and communication research in the Nordic countries (Denmark, Finland, Iceland, Norway, and Sweden). The journal is published in hard copy, but is also available as an open access title on the web. Its aim is to make Nordic media research available to members of the international research community and to promote collaboration between Nordic researchers and colleagues all over the world. Some 1500 people and institutions outside the Nordic countries subscribe to Nordicom Review in hard copy. Distribution is the key factor here. Provided the quality of the contents is good, wide distribution leads to re-publication (possibly in translation) in other journals and anthologies, and representation in widely accessible open access portals and platforms on the web. Nordic researchers in the field of media and communication research have the benefit of a journal that reaches far out into the world — an independent channel that is part of the research community.

Collaboration between disciplines and collaboration across national frontiers, with the aim of enriching the research environment, is vital to the development of fruitful discursive communities. It is also important for research communities themselves to create platforms to achieve long-term goals through national, Nordic and European programmes and projects so as to make their voices heard in the international research community. Open Access can be an important tool toward this end.

But first, a process of what I would call “creative self-examination” is needed — in the Social Sciences and the Humanities alike. We need to take a good look at the relevance of the questions we formulate, we need to develop fruitful theoretical perspectives and to refine our sensibilities regarding choice of methods. We need to consider the ontological and epistemological assumptions embedded in the various methods at our disposal and, not least, we must honestly assess the validity of our findings and the conclusions they support. The outcome of this process will depend on our degree of involvement in discourses outside our institutions and closest circles; that is, on our participation in national, regional and global conferences and networks and international postgraduate exchanges between universities. The importance of networking cannot be overstated.

In this connection, we should be mindful of the importance of a command of English, the lingua franca of our times. We need to have a functional command of English, alongside our native languages in our scientific work. All the evidence suggests that multilingualism stimulates creativity, intellectual versatility and effective communication. I should stress that I mean we need to improve our command of both Swedish and English. Communication skills in these languages should be part of undergraduate and doctoral programmes alike.

How good a measure is publication in international journals?

A separate, but important issue is what measuring stick should we use in assessing the quality of research in the Social Sciences and the Humanities. How fair a measure is publication? Does something go lost, if we concentrate too one-sidedly on international publication? Some quality titles tend toward ever narrower specialisation; some seem inordinately trend-sensitive. The question is, what kind of research will we end up with? Is there a danger of trend-conscious conformity — or of 'Balkanisation', a splintering of the field — as a result? What are the consequences of the fact that what we call 'international publication' today is in essence publication in the Anglo-American sphere? Would the quality of our work perhaps profit from closer contact with colleagues in other cultures close to us, such as the German and French? This, of course, presumes we have a working language of these languages, as well. German and French literature is largely excluded from *Social Science Citation Index*, but each country is big enough to have its own research traditions and conventions.

More and better communication between research communities is of crucial importance if we are to preserve the proper role of Science in an age when responsiveness to external sources of finance ('marketisation') and new objectives for higher education both urge change. Research tends to be more and more administrative, and short-term perspectives prevail — at the expense of the long-term accumulation of knowledge. Again, Open Access has a dual potential to help counteract the pressures toward conformity, and to broaden research horizons through better communication and more frequent exchanges of knowledge.

References

Blomgren, A: *Bedömning bland likar? En jämförande undersökning av peer review-systemets brister i traditionella och elektroniska tidskrifter* [The judgement of one's peers? A comparative study of faults in the system of peer review in traditional and electronic journals], Biblioteks- och informationsvetenskap, Umeå universitet 1997

'Engelska språket som hot och tillgång i Norden' .[English as threat and asset in the Nordic region], Nordiska Ministerrådet, Köpenhamn 2003 (*TemaNord* 2002:561)

Ingwersen, P: 'The International Visibility and Citation Impact of Scandinavian Research Articles in Selected Social Science Field: The Decay of a Myth', *Sociometrics* 49(2000)1, pp. 39-61

Stenius, K: Journal Impact Factor - mittari joka vahvistaa tutkimusmaailman hierarkiaa (An Indicator that Strengthens the Hierarchy in the Research World). Tieteessä Tapahtuu, 21(2003), pp. 35-39

Swales, J: 'Citation Analysis and Discource Analysis', *Applies Linguistics* (1986)7, pp. 39-56

Zuckerman, H. ' Citation Analysis and the Complex Problem of Intellectual Influence', *Sociometrics* (1987)12, pp. 329-338

White, H J: 'Citation Analysis and Discourse Analysis Revisited', *Applied Linguistics* (2004)1, pp. 86-116

Svensk sammanfattning

Hur kan svenska forskare komma till tals i det internationella forskarsamhället? Hur kan ett litet land och ett litet språkområde göra sig gällande. Vad är kvalitet i forskningen inom samhällsvetenskap och humaniora? Inom detta område ryms allt från ämnen som har utvecklat vetenskapliga standards, som kan avläsas i s k kvalitetstidskrifter, medan synen på vad den vetenskapliga standarden är inom andra ämnen i stor utsträckning varierar, vilket avspeglas i tidskriftsfloran. Inom flera humanistiska ämnen är dessutom artikelpublicering mycket begränsad och genrekonventionerna delvis lokala och därmed inte konkurrenskraftiga i internationell mening. Dessa frågor måste diskuteras och föras in i varje forskarutbildning - även om Open Access ger nya och vidare möjlighet till publicering.

En annan central fråga i förlängningen är vilken måttstock som ska användas vid bedömning av samhällsvetenskap och humaniora. Hur rättvisande är internationell publicering som tongivande måttstock? Finns det en fara med att driva frågan om internationell publicering allt för enkelspårigt? Vissa kvalitetstidskrifter riskerar att bli mycket specialiserade och andra mycket trendkänsliga - frågan är då vilken forskning vi får – finns det risk för ensidighet och likriktning. Sådana följder kan bl a Open Access-publicering rätt använd bidra till att motverka.

ScieCom info
Svenskt Resurscentrum
för vetenskaplig kommunikation
www.sciecom.org

Tomorrow's File Endings: On Archiving Principles and Archiving Formats

Marcus Uneson, PhD Student at
Dept. of Linguistics, Lund University
Marcus.Uneson@ling.lu.se

Artikeln i pdf-format för utskrift

Sammanfattning på svenska

Artikeln diskuterar utan tekniska detaljer några viktiga aspekter på långtidsarkivering, särskilt av vetenskapliga publikationer. Efter en översiktlig framställning av fysisk datasäkerhet läggs tonvikten på sunda principer för dokumentarkivering i allmänhet, och särskilt på hur dessa kan vägas in vid val av filformat. Ur den gyllene regeln ”undvik varje val som i onödan begränsar framtida möjligheter” härleds tre grundsatser för arkivering av publikationer: a) använd enkla, textbaserade format; b) använd öppna standarder; c) arkivera en representation, inte en presentation. Några vanliga arkivformat för text idag kommenteras utifrån dessa ståndpunkter, särskilt pdf, pdf-a, tex/latex, xml.

Abstract

The article discusses without technical detail a few important aspects of long-term archiving, in particular of scientific and scholarly publications. After a brief overview of physical data survival, emphasis is laid on general, sound principles for document archiving, and on how such principles might be considered in the choice of archive file formats. From the Golden Rule of Archiving, “do not unnecessarily restrict future options”, three general principles for publication archiving are derived: a)

use simple, text-based formats; b) use open standards; c) archive a representation, not a presentation. With these as point of departure, some common current document formats for archiving are commented, in particular pdf, pdf-a, tex/latex, xml.

Part One: Introduction, Archive sustainability, Physical data survival

Introduction

An increasingly large part of mankind's records is being made available through electronic archives. For the important special case of scholarly and scientific communication, physicists and computer scientists, in particular, have taken leading roles in this process. For many disciplines, most publications can be accessed online; additionally, to a growing extent, this is true also for the many different types of primary data underlying the research results. On the whole, the neural system of scientific and scholarly communication is already electronic.

There certainly remains much to be done to facilitate resource discovery and access. Nevertheless, many researchers have already found that the ease with which they can get at relevant publications within their field has revolutionized the way the do research, and it has done so in a matter of a few years. The results of many millions of research hours are accessible with just a few mouse clicks, or can at least potentially be made so (1).

We have hitherto unparalleled possibilities to climb the shoulders of giants, in all shapes and sizes. Well within the horizon, we might envision a searchable and browsable domain of automatically interlinked publications (for instance, through citation or field similarity). The domain also accommodates the underlying research data, which has been produced and annotated through a world-wide collaborative effort. This domain is an abstraction, for sure; physically, the resources may exist in several copies on many different servers. However, a researcher needn't know, or care: after having identified herself to the system (with a unique, world-wide valid user ID), she may happily romp around much like she

would in her own, local file system.

Such a level of accessibility would itself be enough to baffle researchers of past generations, were they to pay us a visit. Still, we have so far only sketched this virtual domain as a habitat for human researchers. More stirring to the imagination, it will increasingly also be populated by agents – autonomous computer programs which for instance may exploit natural language processing techniques and inference engines to work on text, metadata, and markup. Agents may be employed to classify, index, and automatically ingest new documents into the domain; to summarize documents and provide semantically relevant links between them; to combine information and draw conclusions; and to do many other things we haven't even thought of yet.

Before we get there, there are many problems to tackle, spanning many areas. For instance, on a social level: how do we best encourage and help young and old researchers to learn whatever they will have to learn and to do whatever they will have to do, in order to contribute most efficiently to their respective community? Or economical: what business models for the publishing industry will benefit the research community most, and how can we promote them? Or legal: how do we handle legal matters (e.g., intellectual property issues; ethical use of research data) in a fast changing world-wide distributed environment, where legislation differs between countries and, furthermore, constantly lags behind the technical development?

Certainly, there are technological bites, as well, to chew for quite a while yet (for instance citability, presupposing among other things robust version handling with unique document and version IDs as well as a reliable addressing system; and better searchability, perhaps through more efficient metadata and more mature ontologies). Still, leaving now the more futuristic applications aside, it seems fair to say that the main obstacles for providing, maintaining, and enlarging the basic infrastructure for scientific communication are no longer predominantly technological.

Most technical aspects of this budding research infrastructure are of concern mainly for IT and computer professionals, and there is no need to change that. Nevertheless, there is at least one area where some general knowledge could be useful for most researchers (the 'content providers', as the Newspeak goes), in order to grasp the risks and tradeoffs connected to a certain choice: archive sustainability.

The present article tries to give a bit of non-technical background knowledge of the long-term perspective on document archiving, particularly focusing on the question of archiving file formats for text-based documents, such as typical scientific and scholarly

publications. (The somewhat sloppy term “documents” will here refer to such publications, with no attempt made to find a more precise definition.) However, it doesn’t hurt to have a notion about general sustainability issues either, so we will start one floor up.

Archive sustainability

We can read and with varying degrees of success also interpret ancient media, such as clay tablets, parchment scrolls, and runic stones, thousands of years old. By contrast, mankind hasn’t shown much concern for preserving its more recent, digital heritage for posterity. It is salutary, for instance, to consider that data from the first moon landings are irrevocably lost – even if we could dig up machines that could read the since long obsolete tape format, we would not know how to interpret the undocumented, unstructured bitstream we would find.

We can only strive to do better today. Of course, not all long-term aspects of digital archiving are under the control of the archivist (most archives are funded, set up, and maintained through political decisions, and political situations are known, sub specie aeternitatis, to be transitory). However, where we do have a choice, we should consider the long-term consequences duly.

Admittedly, it is hard to make any detailed predictions about future archiving technologies, in view of the mind-boggling pace of development. Here and now, we cannot think of all problems we might meet in the future, much less solve them. Likewise, we know little of tomorrow’s research methodologies, or their particular requirements. Google will perish, and Citeseer will cease; but the archiving standards we set up should serve their successors, and ours, as well.

In the absence of a crystal ball, we will have to rely on general, sound principles for archiving sustainability. They may provide guidance through uncertainties; they may help us to identify current weak points, and to recognize a better solution when we see it.

A principal challenge of an archive is to cater for the physical survival, interpretability, and usability of the data it holds. Of course, this task is as old as archiving itself; but the dangers and expectations are different in the digital era.

Physical data survival

Storage media deteriorate, at an appalling rate which we have just begun to realize. Indeed, when comparing expected life times of

current data carrier such as CDs (perhaps 20 years), DVDs (possibly significantly less, given the higher information density), or hard discs (typically 5 years or less) to the practically imperishable clay tablets of Sumer, we have little reason to boast about technical progress.

Unfortunately, most research efforts on new media formats goes into increasing access speed and storage space, not longevity. As a result, we get ever faster and more capacious physical formats, with ever shorter life cycles. From a marketing point of view, this is certainly nothing unexpected --- in fact, regularly introducing new formats is a central strategy for customer recycling. Not even ten years ago, small, portable DAT players were state-of-the-art equipment for audio recordings in the field; nowadays, it's practically impossible to find spare parts even if you would dare open one to try to repair it.

Planning for physical survival of digital data under such conditions is not a pleasant task. The only sustainable and scalable solution is to arrange for continuous, automated mass migration: to build an archiving system that is capable of automatically identifying and replacing individual data carriers at risk, and automatically moving the entire archive to a new physical format when the day comes.

This sounds expensive, and it is; but it is also an area where centralization is a very efficient measure – the cost per archived terabyte falls drastically with the size of the archive. National computing centres (or, even better, international ones, thus giving some security against political uncertainties) may offer archiving facilities on. The Australian Partnership for Advanced Computing (2) is an early and interesting step in this direction, with several successors world-wide.

Notes and references

(1) The information compression rate in such a collection is stunning. The world's collected publications in mathematics, for instance, have been calculated to fit on around 100 GB, very soon to be standard on any laptop. Of course, it might be that local copies of that kind will be little needed in an even more networked future, if fast, reliable, and omnipresent connection points will offer access to constantly updated archives.
(2) APAC: **http://www.apac.edu.au/**

Part Two:
Archiving principles and archiving formats

Interpretability and usability -- Archiving principles and formats

How do we choose a good encoding for archiving, i.e., what file format should we use? Obviously, as an absolute minimum, we will want one that we can be reasonably sure we can interpret in the future, even in a future when the software tools used to produce it may be long extinct. There is nothing particularly pessimistic about such a scenario. Tools are dependent on the technology of the day; their development time and their lifetime (a few years, 10-20 at most) are both a twinkling of an eye on archival timescales. The data collections the tools produce or process, on the other hand, have generally required much more resources. Consequently, we will want them to be useful for a much longer time – not seldom indefinitely. This requirement already excludes any format that is not well-documented, open standard, and vendor-independent.

Furthermore, our chosen format should be useful, in the sense that it should support whatever operations we might want to subject it to in the future. This is rather speculative – how could we possibly predict what future generations might want to do with the data, or through what devices they will want to have it presented? We can't, of course; but when choosing digital formats, we can try to observe the very general Golden Rule of Archiving (3): *Do not unnecessarily restrict future options*.

Admittedly, such a wording is too abstract to be of much use in a particular case. However, several other, more specific archival principles can be derived from it. For the special case of archiving text-based scientific and scholarly communication, the following ones are suggested; they are meant to be thought-provoking, rather than exhaustive.

Keep it simple

The World Wide Web has thrived much due to its use of simple, text-oriented network protocols. One computer sends a plain-text request to another one and receives an answer in plain-text. Any decent programmer on any platform can quickly understand the specification well enough to exploit this basic framework, in simple or complex applications. By contrast, employing sophisticated formats means encapsulating data in a shell. If we do, we will need more complex tools to process it; tools that no longer can be written by anyone, tools that will need more maintenance in order to work in changed hardware and software environments; tools that are more likely to contain bugs.

To archive text, nothing is better than text. If we need tagged text, we should use relevant markup, and define new if necessary. Simplicity does not mean lack of expressivity; at most, it might mean a bit of verbosity.

In passing, we might note that text files are easily manipulated,

and that issues around document integrity must be taken seriously. We should not, however, plan for security through obscurity.

Use Open Standards

People who come from the proprietary world and try open source software are often surprised by the fact that programs actually may exchange data through shared file formats – thus, programs may complement each other, rather than compete. (One might find it saddening that there should be something remarkable about this, but such is reality.) You may prefer one tool and I may prefer another, perhaps due to differences in natural disposition or in task at hand; but if we decide on an open file format, our programs may exchange data anyway. The actual programs we use are of no interest and they need not know anything about each other: as long as they both fulfil their part of the bargain, they will exhibit interoperability (as the technical term for this most treasured property goes).

The question of interoperability in an archiving scenario is in principle not very different, only more pressing. A program may need to read a file many years after the file was originally created: by then, there might be no trace left of the creating program, the operating system this program was built for, the hardware it used to run on, or their respective authors (indeed, perhaps not even of the country they used to live in).

Interoperability can only be achieved by strict adherence to a public, non-proprietary, well-documented open standard. Such standards should be designed by truly independent bodies, such as The World Wide Web Consortium (4) – any “de facto standard” may sooner or later turn into a marketing weapon.

Incompatible, secret file formats remain an efficient strategy to forcing clients to do all worshipping at a single altar. However, clients provide the money, and therefore they have a strong bargaining position. There is a growing and gratifying tendency to require from software that it be able to save files into open standard formats -- we can hope that it will be unmarketable otherwise.

The main trap to look out for in this process is that the standards simultaneously are somewhat “improved” (i.e., extended with some arbitrary and redundant features, just enough to make them awkward to use in competing programs). There is a name to this strategy: “embrace, extend, extinguish”. It would not be a serious threat on a balanced market, but given the current situation, where a very small number of vendors are responsible for billions of installations, the danger is real.

Archive a representation, not a presentation

Human inertia is a strong moderator of change rate. Not too long ago, the ‘paperless office’ or even the ‘paperless society’ were envisioned by some writers. As prophecies, both have so far failed miserably; in tech journalist Dick Pountain’s words, IT has rendered paper superfluous in much the same way that the car has made legs unnecessary.

A bit of inertia might be salutary at times. In our child age of digital archiving, our not-so-impressive records of lost data would probably have been even worse, had we been more eager to replace paper by discs. We won’t argue here about the pros and cons of paper. Disregarding environmental aspects, paper is fine, and some future technology might be fine, too (perhaps large, cheap, soft, thin, foldable screens). The point to be made, however, is that there is an unfortunate human tendency to equate a work with its physical presentation; for textual works, this usually means ‘as it is printed on paper’. In the terms of library science, we tend to confuse a work itself with a specific manifestation of a specific expression of that work. For instance, most citation techniques are built around page numbers of a particular edition of a written work, rather than internal references of the text itself.

This view of a work as tied to a particular version with a particular layout printed on a particular page size is problematic, for several reasons. First, it is difficult to sustain in a world where text documents can be presented to the user through a number of different (most of which are yet to be invented), and neither text nor devices are necessarily page-oriented.

Second, and more importantly, such a view is an obstacle whenever we want to use technology for something more than just facilitating paper reading – when we want to go beyond just mimicking current practice. Several of the more visionary applications we have hinted at (and countless others we have not) will be carried out by computer programs, written by language technologists, knowledge engineers, artificial intelligence researchers, and others. Computer programs do not benefit from having to deal with presentation formats – they are not, primarily, paper readers.

A chief hallmark of a good archive document format is that it holds a *representation* rather than a *presentation* (5). Logical, structural, or semantic markup form part of the representation. A specific layout, by contrast, does not; instead, it is generated for a particular presentation in a particular set of circumstances. For instance, in a representation, references are internal to the text, expressed in some dialect of computerese (6). In a particular representation which also happens to be page-oriented, they might instead be converted to page numbers. Taking a slightly more

imaginative example, we can think of texts on history, in which all named entities (persons, places, organizations, etc) have been tagged as such in the representation (manually, automatically, or a combination). Humans usually need no help in identifying named entities, and so the markup need not be seen in a presentation format meant for reading. To information-processing agents, however, such a tagging is of great help.

In fact, it is fruitful to think of a specific presentation format as just another export option from a representation, where all choices can be made according to the needs of that moment: technology available, task at hand, user preferences, presentation device, restrictions of bandwidth, storage space, etc. For a long time to come, one such export option will undoubtedly be printing on paper, but there is no reason to believe that it will be the only one for all future and, above all, there is no reason to choose this particular form for archiving.

Admittedly, it might sometimes be difficult to identify the borderline between content and presentation. For some disciplines, such as legal science, the ability to faithfully reproduce a certain layout may be crucial (see more about pdf-a below). For general publications, however, exact copies are usually not needed (and when they are, a structured representation could be linked to a page image).

Notes and references

(3) This will ring familiar to archivists in general, but perhaps in particular to people working with digitization of cultural heritage. See for instance the Ninch guide,
http://www.nyu.edu/its/humanities/ninchguide/
(4) **http://www.w3c.org/**
(5) These terms can be seen as generalizations of the well-known distinction of “content” versus “style”.
(6) The current tool to do so would be the XML substandard XPath. (It is interesting to consider one of the most widespread and quoted books in existence: the Bible. It has been translated, reprinted, and orally transmitted to the point that nobody would confuse the work with a particular representation. A typical reference may read “Cor. 1:13” – not very different from an XPath expression.)

Part Three:
Current archive format practice, Conclusion

Current archive format practice

The most common formats for text archiving today are native

formats (mostly MS Word), pdf, pdf-a, tex/latex, and different xml applications. These are briefly presented below. Other current formats include sgml (7) (still around for legacy reasons, but unnecessarily complex, little supported and superseded by xml for all practical purposes), html (far too restricted for general use; can easily and better be expressed in xml if needed, as xhtml (8), plain text (rather restricted, but reliable; when nothing more sophisticated is needed, archiving plain text is the closest we can get to digital clay tablets).

Native formats

Software companies are profit-driven (and they can hardly be criticized for being so). Their main responsibility lies with their share-holders; if they choose to publish specifications, or lock them in a cellar, or to continue or discontinue development or support, or to double or halve or setting to zero the price of their products, they do so on approval of their market analysts – to do profit in the short-term perspective (short-term at least from an archivist's point of view).

Native application formats (e.g., MS Word, WordPerfect) are fine for something you know you will never share with anyone, including yourself a few years from now. For any wider or longer perspectives, they are very unsuited. It does not take much fantasy to grasp the risk implicit in locking important data encoded in some binary, closed, proprietary format. There is no guarantee that the data can be recovered at all; if it can, it may cost practically anything. Just to mention a few scenarios: the company behind your program may be put out of business, or discontinue support for your platform or version, or abandon backwards compatibility, or charge ten times more than you expected for the next upgrade, or refuse to fix a bug which happens to be crucial for you.

From an archival point of view, it is important not to be short-sighted (and among other chief virtues of an electronic archivist, we might in particular note parsimony, distrust, and paranoia).

Pdf

The most common choice of archiving format is the portable document format (pdf), created by Adobe. In contrast to most native applications, the pdf specifications up to and including the current has been made public (9), allowing third-party software including some open source projects to create and read pdf. Since the format works well on most current platforms, looks nice on screen, and in particular gives good quality printing for our paper-oriented minds, one might be tempted to think that the question of archiving format is solved.

This is exactly wrong. Adobe controls the pdf format and it may be

changed at any time, with no specification made public. Most pdfs are created by software from Adobe; if the company see fit, it could for instance introduce a new pdf version, “improved” but unfortunately closed and only readable with Adobe software.

More general drawbacks from an archiving point of view is that pdf allows encryption and scripting, both of which should be banned from archiving formats. It also permits embedding of audio and video. (While we certainly should be able to link media files to publications, embedding is not the mechanism.)

Even more generally, the relatively good support for metadata does not change the fact that pdf is strongly presentation-oriented and so less useful for automatic processing. It prints nicely and thereby lets humans go on the way they use to, but it mixes up content and layout and it does not necessarily hold any representation of logical structure.

Pdf-a

Pdf has been criticized for archiving purposes, and rightly so. As a reaction to that, representatives from Adobe and several communities and (US) governmental bodies, especially representing the legal sector, have presented a new, slimmed-down version of the format, known as pdf-a. The format was proposed to ISO as a text archiving standard and recently also accepted.

Pdf-a (a for ‘archiving’) is basically a subset of the pdf 1.4 specification; in particular, it prescribes that all fonts of a document must be included in a pdf-a and that no commercial fonts be used; furthermore, that no encryption, scripts, or embedded media be used. Pdf-a is not owned by Adobe, and several of the objections to pdf are thus bypassed.

Still, pdf-a is just as presentation- (and paper-) oriented as pdf; in fact, it is proposed as a “preferred format for page-oriented textual (or primarily textual) documents when layout and visual characteristics are more significant than logical structure.”(10) When layout is crucial, for instance in court (11), it may form a good complement to other, more structured formats. However, in scientific and scholarly communication, this shouldn’t be too often, once we learn not to confuse work and representation.

TeX

Research communities oriented towards science, mathematics, or computation mostly use the TeX system for communication (12). It is free, stable, extensible, reliable, and does a remarkably good job of typesetting demanding texts, such as mathematical formulae or multilingual works.

A TeX file is made up of pure text, or rather source code, which is compiled into a specific, typeset presentation format. Thus it is enough to archive the source code.

The drawbacks are again its focus on presentation (and exclusively visual presentation, at that). TeX is meant for typesetting, to produce documents to be read from paper or screen by human. Nevertheless, being compiled, TeX sources are strongly syntactically structured; this makes automatic processing much more feasible.

Another drawback is that, even though TeX can excellently perform most of the tasks which are today done by word processors, it is somewhat demanding to use without a bit of technical knowledge. Many researchers have never written a computer program; if so, the thought of directly manipulating source code may be paralysing. Nice graphical user interfaces, such as LyX (13), can possibly reduce the need to do so.

Xml

Xml is a metalanguage, This means that it is a language to design languages, in this case, markup languages. It is very well apt for representation, transmission, and storage of textual information: it is text-based, readable by humans as well as by computers, self-documenting, portable, expressive, international (all xml is in Unicode). In front of all, it is a free, open standard, defined by the W3C Consortium (14).

Among the drawbacks of xml is that it is rather wordy, that it forces data into a hierarchical structure, and that it handles binary data only with difficulty. None of these are very crucial to scientific communication; while xml might not be what we will have for all future, it is about as far as we can get in not restricting future options today.

The particular languages defined in xml are called xml applications, formally specified through particular computer languages known as schemas. An xml document is said to be validated against its schema, thus ensuring interoperability. Furthermore, to the benefit of agents, .xml markup can be combined with a computer-readable semantic specification of the elements, known as rdf.

There are already many xml applications for very diverse purposes. We might for instance note OpenDocument (15), which is a brand-new open file format for general-purpose office uses – but in contrast to proprietary ones, it is an office format we can expect to be able to read in the future. However, to fully exploit the possibilities, communities themselves need to define new xml applications, according to their specific needs – be it for communicating research in generative syntactics or in exospheric

chemistry.

Nobody wants to write xml directly, but given a schema, a generic xml editor or one specifically written for a certain application could be used. However, much remains to be done in terms of user-friendliness – editing xml is currently no more pleasant than producing TeX.

Conclusion. Now what?

Survival of digital data will be a growing concern in all corners of society, and any single measure will be hopelessly insufficient. Still, a well-known prescription is to offer tools and education, and the world of digital scientific communication is no exception.

Although far from optimal, we will have to live with pdf as archiving format for some time to come (pdf-a perhaps for quite some time). Likewise, the far better TeX format will live long, together with xml (automated conversion between the two is nothing impossible.)

From a text-archival point of view, however, it seems clear that xml currently is the best choice for long-term purposes. However, to make xml useful, we need good tools to help in writing – general xml tools, community-specific xml tools, word processors, any tool is fine, as long as it might be persuaded to produce valid xml according to the community's schema. But first we need standards – xml and rdf schemas specified by the respective communities and suited to their particular wishes.

Education is another important point; on a general level, all researchers should know about the possibilities of digital communication through space and time; but also about the risks involved. More specifically, although most researchers are not too interested in technical details of file formats (and they shouldn't need to be), they do care about the survival and usability of their work. A tiny bit of digital long-term hygiene would not be out of the way in any curriculum (and strongly recommended also for senior researchers).

Notes and references

(7) see for instance **http://www.w3.org/MarkUp/SGML/** or **http://xml.coverpages.org/sgml.html**
(8) **http://www.w3.org/TR/xhtml1/**
(9) **http://partners.adobe.com/public/developer/pdf/index_reference.html**
(10) **http://www.digitalpreservation.gov/formats/fdd/fdd000125.shtml**
(11) see for instance **http://www.scientific-computing.com/**

scwmayjun05archive.html
(12) A good introduction is found on http://www.ctan.org/what_is_tex.html
(13) http://www.lyx.org/
(14) http://www.w3.org/XML/
(15) see for instance http://www.oasis-open.org/news/oasis_news_05_23_05.php

ScieCom info
Svenskt Resurscentrum
för vetenskaplig kommunikation
www.sciecom.org

How do we motivate our researchers?

Inge-Bert Täljedal, Vice-Chancellor,
Umeå University
rektor@adm.umu.se

Presented with the question how researchers could be motivated to publish for open access, my first reaction is simply to wonder why we should at all bother. The advantages of open access arise more or less automatically from the historically unique properties of the web which can be clearly seen by young people, the academic leaders of to-morrow. So, it might be a bit presumptuous to believe that librarians or university presidents must do something special to motivate the researchers.

However, let us for the sake of the argument accept that for some short time still, it is not an expression of redundant vanity to try and open the eyes of our researchers to the new possibilities. How should we then go about it?

First of all, missionaries must be convinced that the new alternative represents a step forward with respect to the goals of science itself. Relying on non-scientific arguments in favour of digital publication could at best be of indirect value. For example, economic arguments would presumably count as relevant only insofar as the money saved would be at the disposal of the scientists themselves. Merely to point out that digital publication could save money for the university systems of the world, or for the local library, would probably have little impact on the individual researcher.

Like people in general, scientists are a heterogeneous lot with regard to psychological make-up and professional aspirations.

Some are driven by a strong curiosity and desire to know genuine truths about the world. Others, usually in the humanities and social sciences, seem equally keen on denying objective truth as a regulatory idea for research. Some are spurred on by a desire to become famous, while others have more relaxed goals in life. It seems possible that an internet-based system of archives and digital journals, with or without open access, will appeal differently to different personalities. For the purpose of our question, it seems acceptable to focus on an imagined group of academic trend-setters, recognizing that this is indeed a simplifying construction. What would motivate this model researcher to publish his or her reports for open access, or at least electronically?

A special aspect of scientific culture should be borne in mind, viz. the constant dialectic tension between strong conservatism and an equally strong revolutionary longing. Many logical and empirical pitfalls lie ahead of every new research project, a fact which makes it mandatory to safe-guard rigor and stringency in methodology. Whoever proposes something really new should therefore be suspected of having overlooked some source of error. Science recognizes that we are constantly making errors and only occasionally great discoveries. Hence, good scientists are people with strongly conservative tendencies.

At the same time, original ideas are highly praised, provided they are not too easily killed by criticism. There can be no greater achievement than finding out something completely new about the world, a result that is both unexpected according to established thinking and of wide-spread consequences. Hence, good scientists have revolutionary aspirations.

This constant balancing of conservatism against radicalism is a subtle challenge. The not so good researcher can fail in either direction. Some inhibited people are too afraid of making mistakes and always prefer to be on the safe side of the established methods and theories. Their results are of limited significance because of limited originality. Others are too uninhibited and care too little about the methodological norms. They may have visions but their contributions are slim because they are too speculative and unreliable. The successful researcher knows when and in what sense to be brave and chance-taking, and when to play it safe.

It is the right balance between the cultural heritage and the unexpected new insights and new ways of doing things, the optimum blend of conservatism and radicalism that constitutes quality in the academic world. High-quality research strives for genuinely new insights but does so from a venerated base of ideas, methods and norms of conduct. Publishing a research report is an integral part of the research process itself. Therefore, if electronic publication were to be seen by scientists as a token of relevant

modernity, in contrast to the methodologically outmoded publication in print, that perception would probably constitute a strong motive force in favour of e-journals and self-archiving.

That the mode of publication matters a great deal to researchers reflects a fundamental recognition of the fact that Truth is evasive and difficult to come by. To realists, who believe in objective truth, a refined form of international cooperation is necessary, a social system of constructive criticisms that requires established methods – not only for experiments but for the exchange of ideas as well. Metaphysical non-realists and constructivists may have other, but no less compelling, reasons for viewing research a collective social enterprise. If Truth is not objective but subject to negotiation, some agreement on the rules of negotiation, i.e. on the rules of publication, becomes mandatory.

As the mode of publication is an integral aspect of the quality concept, it is rather a delicate question how one could justify a change of publication habits. I think it is fair to say that the academic world strikingly resembles those of sport or entertainment in shamelessly appreciating fame and social success. This fact must be recognized as such, whatever we think of the dictum by Erich Fromm, the famous psychoanalyst who asserted that wanting to become famous is a sign of insanity. Ideally, research reports should confer recognition in proportion to their quality. Therefore, a general answer to the question raised could be as follows: ‘To motivate researchers to publish in open access journals or archives, one should demonstrate to them that such a mode of publication affords a higher quality to the report than traditional publication, or at least signifies that the report is of unusually high quality.’

Science requires that researchers can communicate easily with each other. As the learned communities are expanding world-wide, and as the literature potentially relevant for anyone researcher is also increasing, the easier dissemination of information by the web than by paper and ink clearly speaks in favour of electronic publication. In the same vein, the relative ease by which information on the web can be retrieved must also be considered an objective advantage. Not so long ago a forefront scientist could more or less know by person all the people with whom to interact in the international discussion. This is no longer so, except for very narrow fields of inquiry.

Focusing more on the intrinsic properties of a scientific report, the fact that the web lends itself to hypertext technologies is interesting. A hundred years or so ago, the first automobiles were designed as horseless horse carriages. Similarly, it is natural first to think of the digital report as a printed paper without print. But the process of reshaping the traditional research report into a

modern Saab is already underway. Those interested in producing high quality science reports will have to take that into account. A few times I have had reasons to cite papers a hundred years or so of age and have been fascinated by their poverty of documentary and illustrative material. The young scientists of to-morrow will probably look with similar fascination on the format poverty of the printed papers of to-day.

When confronted with the technical options for publication on the web, many researchers seem to be afraid of wasting their good results on low-status archives. They fear that their reports will not be rated at the high scientific level that their intrinsic qualities justify. There are two realities behind this kind of concern. One has to do with the role of peer reviews in the scientific world. The other stems from a modern decline of the principles for quality assessment in connection with the distribution of grants or the hiring of academic staff.

The system of peer review is so well established that many scientists regard it a criterion of scientific quality and tend to publish in peer-reviewed journals only. However, the system has dual effects. It both weeds out low quality manuscripts and makes it difficult for highly original papers to get published. Peer review stimulates the production of main-stream, medium quality work.

Nonetheless, the usefulness of peer review is such that it is impossible to forecast great success for any web-based journal which does not incorporate peer review. Repositories, entirely open to authors and readers, could function as complementary elements in the global system of electronic publication. Access to open searchable archives, such as those established by the universities of to-morrow, would in fact mean a safe-guard against the risk that the peer-reviewed electronic journals, by mistake or ignorance, suppress any intrinsically valuable report just because it does not match the received modes of thought for the time being. A self-archived good report need not sink into oblivion as long as it can be referred to in other reports published in a peer-reviewed journal. Thus, self-archiving in what might perhaps be frowned upon as dull, indiscriminate repositories or dumps need not at all be a bad thing for the advanced researcher – as long as there are well edited e-journals as well.

However, researchers may worry that publishing in a not very prestigious context is little appreciated by grant committees and staff recruiting bodies. Such worry is not irrational but quite understandable, considering the way academic culture has developed. Traditionally, a scrutinizing expert was assumed to read the candidate’s papers and books and evaluate their intrinsic quality. As scientific publishing has grown, it has become increasingly demanding to base one’s judgment on thorough

reading. The temptation to rely on indirect markers, such as the locus of publication, becomes strong and sometimes irresistible to some people. It is not uncommon for modern experts to express themselves along such indirect lines, making reference to journal impact factors for example. Such conduct represents a deterioration of a most fundamental academic role, that of the scientific or scholarly expert.

Although I resent and deplore the use of citation and impact numbers in situations where comments on the scientific substance matter would be appropriate, I do not wish to ban the indirect quality criteria from all contexts. The literature is so vast that selective reading is necessary. Citation and impact numbers have a role to play as heuristic tools in a rational literature selection process. But that is something quite different from being an ingredient in the very concept of quality.

The freedom to express whatever opinion in a forum of one's own choice is an important aspect of academic life that safe-guards the integrity and moral independence of the university researcher. To infringe upon this basic freedom merely to promote the transition from publication in print to publication on the net cannot be justified. However, legislators and university boards would do well in deciding that all reports that have been published elsewhere should also be deposited for open access, as soon as the relevant copyrights permit. For some time ahead, the copyrights will vary between reports depending on where they were first published. This lack of uniformity is no good excuse for delaying the introduction of routines for as much open access as possible, without infringing on the rights of the authors to decide on the place of first publication.

In brief, librarians and other technical experts should cheerfully go on developing the tools for electronic publication and open access. This work should be supported and encouraged as an investment by the national and local leaders of the systems for higher education. The good objective has little or nothing to gain from imposing administrative rules on the researchers in order to bribe or punish them. Instead, they should be shown how the modern tools for publication are in fact advantageous to their own basic ambitions to be up-to-date, forward-looking, and keen on preparing as good and striking research reports as possible.

Svensk sammanfattning

Forskare i allmänhet eftersträvar erkännande för kvaliteten i sin forskning. Att publicera är en del av forskningsprocessen. Publicering för allmän tillgänglighet (open access) via nätet framstår som en attraktiv möjlighet endast om tekniken kan uppfattas öka arbetenas kvalitet. Kvalitetsfaktorer av betydelse är:

effektiv spridning, enkel sökbarhet, möjligheten till hypertext. Forskare bör inte tvingas att övergå från konventionell till digital publicering. Ett rimligt krav är däremot att publicerade arbeten ska deponeras i ett OA-arkiv så snart copyright-reglerna tillåter det. Dessa idéer diskuteras i relation till psykologiska aspekter på forskares beteende, impactsiffrors relevans och kvaliteten i sakkunnigas utlåtanden.

ScieCom info
Svenskt Resurscentrum
för vetenskaplig kommunikation
www.sciecom.org

DSpace Federation 2nd User Group Meeting

Jonas Gilbert, Web Coordinator for
the Digital Library, Göteborg University
Jonas.Gilbert@ub.gu.se

One of the most used tools world wide for Institutional Repositories is the DSpace software. It is today believed to be the most used platform for IR in the US, while EPrints still is the commonest platform in Europe. DSpace was developed by MIT and Hewlett-Packard and version 1.0 was released in November 2002 under an open source-license. While the initial development of the system was funded by MIT and HP, the subsequent development has depended on the efforts of the user community. The R&D section of Hewlett-Packard – HP Labs – still keeps a close engagement in the project and the system architect for DSpace, Robert Tansley, holds a position there. HP looks at this investment as a way to be leading in the area of digital preservation, and says that they will stay involved as long as the community shows a keen interest for the platform. MIT, as a user of DSpace, obviously has an interest in the progress of the system and the staff that works with DSpace at MIT has been very important to the whole user community by their efforts to coordinate websites, administer mailing lists and arranging user meetings.

A first user meeting was held in Boston in March 2004, and a second user meeting was held July 7-8 this summer in Cambridge, UK, and gathered 140 participants from 22 countries. A recurring theme in the presentations and discussions during the conference was the future development of the system, both from technical as well as functional and organizational aspects. Presentations were given by among others Clifford Lynch, Executive Director of the Coalition for Networked Information and Matthew Cockerill,

Director of Operations of BioMed Central.

While DSpace originally was designed as an IR platform for academic communities, there are today many examples showing how the software is employed in a wider variety of purposes. DSpace is for instance being used in connection to an extensive Chinese digitalization project concerning museum collections held at Chinese universities – in all more than 100 collections each calculated to require 2 terabyte data when digitalized. Another example, and from a totally different area, is found in Kansas where the federal government is using DSpace as a repository for the state documentation. Yet another example shows DSpace as a tool for e-book publishing, something which has been tried by The Australian National University. In India the authority responsible for the collection of statistical data is using DSpace as a storage tool. In UK DSpace is being used to e-publish the reports from a national research program within the social sciences, and in connection to this project the administrators have developed visualization tools for DSpace that allow for analyses of how the researchers from different universities are collaborating. This list of how DSpace is being used in different areas could be made longer, but the main area of use is of course in line with the original purpose for the system, i e as an Institutional Repository for universities. A question that was raised at the user meeting concerned how these different areas could be made to collaborate in the further development of DSpace? Should the thousand flowers bloom, or is it of importance that the system has a clearly defined roadmap?

To address this question about the future roadmap for DSpace, MIT and HP will appoint an interim board whose object will be to consider the future administration and organization for DSpace. A report from the interim board is due next spring, and one main issue is to describe the position of DSpace with regard to the broader development within digital preservation and e-publishing. This question was also addressed by the closing speaker Clifford Lynch, who pointed out that DSpace of course not is developed in isolation, but also that we today don't fully comprehend the environment and that we don't know how a repository system such as DSpace in the future will interact with e-learning platforms, publishing tools and other applications. Lynch also observed that Institutional Repositories today is a serious question on the agenda for many research institutions, and that approximately 80-85 % of all universities are implementing or planning to implement an IR. Along with this Lynch also noted that research funding bodies such as NIH increasingly are demanding that applications for research projects should include plans for data management.

Lynch also said that it is possible to see a difference between Europe and US with regard to how the IR systems are being implemented and used. In Europe the e-publishing of the

traditionally printed documents is in focus, while the universities in US are more interested in integrating other types of material in the repositories. This could be material such as research data, video or sound recordings, course catalogs, newsletter or other kinds of documentation related to education and research. Lynch did not consider either of these ways to use the repository as a “wrong” way, but said that we must keep in mind that the way the system is being used naturally will guide the strategies for the future development. Should an IR, for instance, be able to manipulate the data? Or should all items be just “big bags of bits” that leaves all processing of the data to the user? There are also several important issues left to solve concerning the long term preservation. For instance must a way to geographically distribute the archives by mirroring be achieved. Lynch also concluded that the best guarantee for an archive to be operational in the future is when it is being used daily now.

BioMed Central is using DSpace for their service Openrepository.com, which is an opportunity for organizations not willing to install and run a repository on their own. BioMed Central works as the service provider, hosting the repository and offering different related services for the customers. For this purpose BioMed Central has been active in developing new functionality for DSpace. In his presentation Matthew Cockerill from BioMed Central declared that these developments would be made accessible to the DSpace community.

A Nordic contribution on the user meeting came from the University of Bergen and librarian Elin Stangeland who described the plans for making a connection between the DSpace repository in Bergen and the Norwegian national system for research documentation, FRIDA.

It is very stimulating to follow the discussions about the future development of the DSpace platform. It is a good example of the possibilities and issues that comes with an open source project. If there sometimes in the discussions were a focus on problems, others would point out that these are natural questions that follow with a successful project. The need to co-ordinate the development of the system and the information and support is obvious. There also was a discussion about what it means to be a registered user of DSpace, and if there should be a more formalized possibility to enter the community of DSpace users. Many current DSpace users are already giving important contributions to the development. Among the examples of functionality that has been first locally developed and then integrated in the main software are: multi language support (University of Patras, Greece), authority lists for subject headings (University of Minho, Portugal), possibility for the readers to write comments to items in the archive (New York State University)… this list could be made longer and is a good illustration of the benefits that the users in the community can

have from each other in an open source project.

Svensk sammanfattning

Ett av de mest spridda programmen för Institutional Repositories är DSpace, utvecklat av MIT och Hewlett-Packard och lanserat i november 2002. I dag beräknas fler än 100 installationer av DSpace finnas i drift, spridda över hela världen. Många av dessa användare deltar nu aktivt i utvecklingen av DSpace, genom att skapa nya funktioner som sedan integreras i själva programvaran. Ett användarmöte hölls i somras i Cambridge och samlade 140 deltagare från 22 olika länder. Vid mötet diskuterades frågor kring den fortsatta förvaltningen av DSpace, och MIT och HP presenterade en plan för att tillsätta en interimstyrelse med uppdrag att utreda hur det framtida samarbetet kring DSpace ska organiseras. På användarmötet talade bland andra Clifford Lynch från Coalition for Networked Information som konstaterade att Institutional Repositories är en viktig fråga på många lärosätens agenda, och att 80-85 % av universiteten idag antingen driver eller planerar att driva ett institutionellt arkiv. Samtidigt blir det också allt vanligare att forskningsfinansiärerna ställer krav på att ansökningar ska innehålla planer för data management.

ScieCom info
Svenskt Resurscentrum
för vetenskaplig kommunikation
www.sciecom.org

Debatt - Lars Klasén, Erik Borälv

Lars Klasén, Informatiker, Infodata AB
lars.klasen@infodata.se

Foto: Renato Tan

Erik Borälv argumenterar i sin artikel **"Fri forskning, men ej fri tillgång till resultaten"** (ScieCom Info 2005:1) för Open Access. Tyvärr gör han det genom att delvis förvränga verkligheten.

Erik får det att låta som om de kommersiella förlagen hindrar forskningsresultat att bli fritt tillgängliga. Det är inte sant. För det första står det fritt för forskarna och/eller deras uppdragsgivare att välja var de vill publicera resultaten. Och även om de väljer de kommersiella förlagen så kan de - om de vill - ge sina resultat spridning via en mängd andra kanaler; något som också de allra flesta gör. Bland dessa kanaler finns rapporter, konferenser, pressreleaser, sammandrag på webben, intervjuer i media, etc.

För det andra så har de största förläggarna av vetenskapliga tidskrifter Elsevier och Springer, liksom ett flertal andra, en policy som tillåter forskarna publicera innehållet i sina artiklar på sina egna institutioners webbplatser, arkiv etc. Detta nämner Erik inte med ett ord - trots att han självklart är bekant med det.

Erik påstår vidare att författaren "måste" ge upp upphovsrätten till sitt manus. Han berättar dock inte att det inte gäller alla tidskrifter och inte för evigt. Och han berättar inte att förlagen samtidigt gör de författare som vill skydda sina alster en tjänst genom att bevaka upphovsrätten till deras artiklar.

Han påstår också att författaren "skänker" bort sitt manus. Skänker? Nej, förfarandet innebär ett byte av tjänster, där förlagen bistår med granskning, bearbetning, publicering, spridning, marknadsföring, bevakande av upphovsrätt, mm. Och i vissa fall honorar.

Så förlagen gör givetvis nytta. Att de inte gör detta av altruistiska skäl är självklart. Om det är förlagens vinster som Erik reagerar på så borde han skriva det i klartext. Erik nämner vinstmarginalen 40 % för förlagens publicering av vetenskapliga tidskrifter. Undrar om han tror på det själv. Den refererade rapporten är två år gammal, de uppgifter den baseras på ännu äldre - och mycket har hänt sedan dess. Och som sagt: det står författarna fritt att publicera sig via förlagen eller inte. Precis som det står förlagen fritt att välja publicering eller inte.

Erik tar upp förfarandet med att betala för publicering. Han nämner dock inte vad det kostar att köpa sig en plats i en tidskrift. Det varierar, men storleksordningen är vanligen 15.000 - 25.000 kr. Det är intressant att sätta detta i relation till de "9000 kr per forskare och år" som han räknat ut att Uppsalaforskarna betalar för att få tillgång till "12435 e-tidskrifter" plus de pappersbaserade.

Förresten är det långt ifrån alla de 12435 e-tidskrifterna resp. pappersbaserade som kan betraktas som vetenskapliga tidskrifter! Så även i det avseendet förvränger Erik verkligheten.

Erik Borälv, Forskare, Nationellt IT-användarcentrum (Nita), Uppsala
Erik.Boralv@it.uu.se

Det ser ut som om jag misslyckades med att föra ut mitt budskap i artikeln "Fri forskning, men ej fri tillgång till resultaten" [1]. Lars Klasén läser in något för mig oväntat, att jag tror kommersiella förlag inte gör ett bra arbete eller rent utav vill hindra spridning av forskningsresultat. Nu är det inte så.

Vad jag speciellt ville peka på var att det i forskarvärlden idag finns ett växande missnöje med nuvarande publiceringssystem. Publiceringssystemet är samtidigt, i praktiken, den viktigaste delen

av det akademiska meriteringssystemet. Det är anledningen till att det missnöjet spelar viss roll för oss forskare.

Invändningarna som jag hade hoppats skulle gå fram var: (i) publiceringssystemets tröghet (traditionell publicering tar för lång tid), (ii) otillräcklig egen kontroll (copyright och tillgång till artiklar) och (iii) fel ekonomisk modell (höga priser, samt att tidskrifterna oftast köps tillbaka av samma personer som publicerar sig) [2]. Dessa invändningar är mer av principiell karaktär och handlar bara till viss del om nivåer eller möjligheter till undantag. Det är därför, till exempel, i sammanhanget inte så värdefullt att man idag kan begära undantag från förlagens copyright, och eventuellt få formell tillåtelse att publicera sin egen text i annat sammanhang.

Lars antyder att det hänt något med publiceringssystemets priser de senaste två åren. Oklart vad eller varför? I mitt enkla räkneexempel försökte jag ge en ögonblicksbild av hur det ser ut om man är forskare på Uppsala universitet och genom sina anslag och forskningsmedel betalar för tillgången till bibliotekets alla resurser. Att det aldrig kan bli en exakt uträkning av hur mycket prenumerationerna kostar per forskare kan inte spela någon roll. Att som Lars säga att en sådan uträkning är en förvrängning av verkligheten och istället slänga in en ny kostnad (tagen ur luften?) om 15.000–25.000 kronor per artikel – vad betyder det? Egentligen spelar det för debatten ingen roll hur mycket publicering kostar; oavsett pris så måste kostnaden betalas. Frågan är hur och av vem. Jag misstänker att det är här vi tycker som mest olika. Jag vill ändå fortsatt hävda att det finns rimliga möjligheter för andra publiceringslösningar att ge en totalt sett billigare lösning än vad de kommersiella förlagen kan ge.

Lars säger helt korrekt att forskare är fria att väja hur de vill publicera sig. Det är dock en fråga som är mer komplex än ett enkelt val, då det hänger ihop med den egna meriteringen. Det är något som akademin själv måste lösa – det har egentligen inget med förlagen att göra. Vad Lars missar är att en del av de alternativa publiceringssätt han räknar upp är olika sätt att gå runt problemet med den traditionella publiceringen.
Jag vill påstå att detta fria val har börjat gå hem. Forskare, universitet – och framför allt bibliotekarier – har insett att det finns en möjlighet till grundläggande förändring här. Vi ser varje vecka exempel världen över på att universitet och finansiärer kräver en övergång till Open Access (OA) eller andra liknande lösningar [3, 4, 5, 6].

Jag håller med om Lars beskrivning om att forskare och förlag byter tjänster med varandra, och att förlagen ser till att utföra många tjänster åt forskarna. Jag har inte alls problem med att förlaget tjänar pengar på det. Dock, jag menar att det går hitta nya lösningar för publicering som är både bättre och effektivare än

(går-)dagens [7]. Bättre i meningen att tillgång till forskningsresultaten blir fri för alla, och effektivare i meningen att en bättre totalekonomi ger möjligheter till mer forskning för samma summa pengar.
De många positiva exemplen med OA som vi läser om är dock ingen given eller enkel lösning. Det behövs först och främst stöd från universiteten så att detta blir en accepterad publiceringskanal. Finansiärernas stöd skulle också betyda mycket då det skulle ge ett snabbt genomslag. Viktigast av allt är att även en medvetenhet om OA når högre än dagens, ska vi säga, gräsrotsnivå. Som jag tidigare poängterade, det finns inga som helst spår av detta på den högsta politiska nivån. Det är synd.

[1] **http://www.sciecom.org/sciecominfo/artiklar/boralv_05_01.shtml**
[2] **http://www.library.yale.edu/~llicense/ListArchives/0410/msg00093.html**
[3] **http://www.theregister.co.uk/2005/05/11/open_access_research/**
[4] **http://www.rcuk.ac.uk/press/20050628openaccess.asp**
[5] **http://segate.sunet.se/cgi-bin/wa?A2=ind0504&L=biblist&F=&S=&P=20916**
[6] **http://www.doaj.org/**
[7] **http://www.dlib.org/dlib/june04/harnad/06harnad.html**

ScieCom info
Svenskt Resurscentrum
för vetenskaplig kommunikation
www.sciecom.org

SUHF går vidare med Berlindeklarationen

SUHF:s styrelse beslöt vid sitt möte 8 juni att rekommendera medlemmarna att vidta följande åtgärder i syfte att förverkliga Berlin-deklarationen:
1. Införa en policy som starkt rekommenderar att deras forskare deponerar en kopia av varje publicerad artikel i ett öppet, digitalt arkiv och
2. Uppmuntra forskarna att publicera sina forskningsartiklar i fritt tillgängliga vetenskapliga tidskrifter när en lämplig sådan existerar och ge det stöd som krävs för att detta ska vara möjligt.

Se SUHF:s hemsida
http://www.suhf.se/Main.aspx?ObjectID=213

Ny milstolpe för DOAJ

2005-08-25 kom Directory of Open Access Journals upp i 1700 kvalitetskontrollerade titlar **http://www.doaj.org/**

Striden om PubChem

ACS - Amercan Chemical Society har gjort stora ansträngningar för att hindra NIH:s (National Institutes of Health) fritt tillgängliga databas PubChem från att utvecklas. Trots intensiv lobbyverksamhet lyckades de inte utan Representanthuset har uttalade sitt stöd för NIH i deras fortsatta arbete med PubChem och ber endast NIH att samarbeta med den privata sektorn för att undvika onödig duplicering. PubChem har en årlig budget på USD 3M och CAS Registry (som ACS vill skydda) har ca USD 260M och en personal på 1300 mot PubChems 13. ACS har på senare tid fått kraftig kritik från sina medlemmar för de mycket höga ersättningar de betalt till sin ledning. Som non-profitorganisation är ACS skattebefriad. Läs mer i **US Congress fails to back ACS**, *Information World Review*, 2005-06-16 och i det kritiska brev som University of California Academic Council skrivit till ACS. **http://**

www.universityofcalifornia.edu/senate/committees/acs.0605.pdf

Författares attityder till egenarkivering

Alma Swan och Sheridan Brown publicerade i maj i år Open access self-archiving: An author study, http://cogprints.org/4385/01/jisc2.pdf.

Studien visar att nästan hälften av de tillfrågade forskarna (49%) egenarkiverat åtminstone en artikel.de senaste tre åren. Postprints deponeras oftare än preprints. De mest produktiva författarna egenarkiverar mer. Fortfarande är många (36%) omedvetna om möjligheten till egenarkivering Bara 10% kände till SHERPA/RoMEO listan över förlagspolicies ang. pre- och postpublicering. 72% använde Google för att söka efter vetenskapliga artiklar.

Ett viktigt resultat av studien är att en stor majoritet - 81% - villigt skulle egenarkivera om deras arbetsgivare eller anslagsgivare begär det. Ytterligare 13% skulle också göra det om än motvilligt. Endast 5% skulle avstå.

Inledningsvis ger författarna den intressanta upplysningen att varken **American Physical Society (APS)** eller **Institute of Physics Publishing (IOPP)** noterat någon förlust av prenumerationer trots att de samexisterat med det öppna fysikarkivet **arXiv** i 14 år. **arXiv** ses inte som ett hot, snarare tvärtom.

Tyska forskningsrådets översikt över attityder gentemot OA

Deutsche Forschungsgemeinschaft har nyligen publicerat en stor studie av OA, Publikationsstrategien im Wandel? Ergebnisse einer Umfrage zum
Publikations- und Rezeptionsverhalten unter besonderer Berücksichtigung
von Open Access. http://www.dfg.de/dfg_im_profil/zahlen_und_fakten/

Över 1000 tyska forskare inom alla discipliner har tillfrågats om sina erfarenheter av OA tidskrifter, och OA arkiv. Svaren visar starkt stöd för OA. Över två tredjedelar svarade t ex att OA kommer att förbättra tillgången till vetenskaplig kunskap och medföra bestående förändringar av det vetenskapliga publikationslandskapet.Trots det uttalat starka stödet var det hittills få som aktivt publicerat i OA, bara var tionde hade t.ex publicerat i en OA-tidskrift.

Inom STM-området var de yngres stöd för OA något starkare än

stödet hosde mer etablerade medan motsatt förhållande rådde inom humaniora och samhällsvetenskap. STM-forskare är mer beredda att betala för publicering än H och S.

På frågan om hur DFG kan främja OA föreslog man att DFG skulle intensifiera debatten om fritt tillgängliga publikationer, försäkra sig om kvaliteten hos OA-tidskrifter samt ge tekniskt, juridiskt och organisatoriskt stöd till öppna postprint-arkiv.

Open Access Law Program

Creative Commons / Science Commons startade i början av juni ett nytt program för att göra juridisk forskning fritt tillgänglig online utan onödiga copyright- och licensrestriktioner. Programmet samarbetar med ett stort antal juridiska tidskrifter för att uppmuntra dem att antingen själva OA-arkivera de artiklar de publicerar eller tillåta författarna att göra det. För att främja OA inom juridiken har Science Commons tagit fram resurser som t.ex. *Open Access Law Journal Principles och Open Access Law Model Publication Agreement*. **http://creativecommons.org/press-releases/entry/5464**

Nyheter från PloS

Den 24 juni lanserades. **PLoS Computational Biology** i samarbete med
International Society for Computational Biology (ISCB). Chefredaktör är Philip E. Bourne, professor vid Department of Pharmacology vid University of California San Diego, co-director för Protein Data Bank och senior rådgivare inom Life Sciences vid San Diego Supercomputer Center.

Grundarna Philip E. Bourne, Steven E. Brenner och Michael B. Eisen förklarar sin vision: "Open access—free availability and unrestricted use --to all articles published in the journal is central to the mission of PLoS Computational Biology, and distinguishes this new journal from most scientific journals which still needlessly restrict access to their contents. Open access revolutionizes the way we use research literature, and takes much inspiration from the field of computational biology itself. “

Den 25 juli lanserades **PLoS Genetics.** Chefredaktör är Dr. Wayne N. Frankel, senior forskare vid The Jackson Laboratory i Bar Harbor, Maine. Redaktionens budskap "Genetics and genomics research have lead the way for timely, open access policies to all types of biological data—it is high time that we applied the same principle to our papers and unleash our creativity to develop new ways to use the scientific literature,"

Den 30 september lanseras **PLoS Pathogens**, men redan nu kan man få en förhandstitt, Chefredaktör är John A.T. Young, professor

vid Infectious Disease Laboratory at the Salk Institute for Biological Studies. Tidskriften kommer att publicera rigoröst bedömda artiklar inom hela fältet av patogener inkl.bakterier, svampar, parasitee, prioner och virus. "Our open-access license means [the research published] is immediately available to scientists all over the world," sager redaktionen.

PLoS Biology har nu en hög impact factor

Den 27 juni tillkännagavs att ISI gett PLos Biology en impact factor på 13,9, som placerar den bland de mest citerade tidskrifterna inom livsvetenskaperna, trots att den ännu inte är två år gammal. **http://www.plos.org/news/announce_pbioif.html**

Stigande impact för BioMed Centrals OA-tidskrifter

Enligt ISI har nu BMC tidskrifter impact factors som väl hävdar sig in konkurrensen med prenumerationstidskrifter Att tidskrifter som bara har ett par år på nacken så snabbt kan få goda impact factors visar att impact och OA inte står i motsättning.

Några exempel: *Arthritis Research & Therapy* gick från 3,44 till 5,03 och rankas nu som nummer två inom reumatologi. *Respiratory Research* fick 5,53och hamnade omdelbart på andra plats inom fältet.

RAE vill minska betydelsen av prestigetidskrifter

Enligt uppgifter i *Times Higher Education Supplement*, 22 juli, 2005 har ledande akademiker involverade i 2008 års RAE – Research Assessment Exercise (mycket stora regelbundna utvärdering av universiteten i UK) uppmanat universiteten att upphöra med sin fixering vid märkestidskrifter som *Nature* och *Science*.
Sir John Beringer, ordförande i Panel D för biologiska vetenskaperna sager: "The jolt will come for those (academics) who take the mindless approach - 'I have so many publications in journals X and Y, therefore I am excellent'. It is terribly important to break the link that publishing in a journal such as Nature is necessarily a measure of excellence."

RCUK, Research councils UK, publicerar förslag till policy om OA

Engelska regeringens påpekade i sitt rätt kallsinniga svar på rekommendationerna i underhuskommittens rapport**Scienctific publications free for all**? **http://www.publications.parliament.uk/pa/cm200304/cmselect/cmsctech/399/39902.htm**
att de offentligfinansierade engelska forskningsråden (åtta stycken inom paraplyorganisationen RCUK) var obundna av regeringens

ställningstagande och var fria att besluta om eventuella krav rörande den forskning de finansierar.

RCUK tog upp den kastade handsken och publicerade i somras sitt **Position Statement on Access to Research Output** **http://www.rcuk.ac.uk/access/index.asp**, ett förslag där grundtankarna från Berlindeklarationen och underhusets rapport kommer igen. RCUK kräver bl,a att kopior av artiklar baserade på rådsfinansierad forskning snarast möjligt skall deponeras i ett öppet arkiv om licens- och copyrightbestämmelser så tillåter.
RCUKs förslag gäller nya anslag från 1 oktober i år, men tidigare anslagsmottagare kommer att uppmuntras till (men inte avkrävas) deposition. RCUK kommer att tillse att anslagsökande får inkludera kostnaden för publicering i OA-tidskifter i sina ansöknngar.

*The Association of Learned and Professional Society Publishers (**ALPSP**)* har skrivit ett kritiskt svar på förslaget och befarar katastrofala följder för sina medlemmar. De får i sin tur svar på tal i ett **Open Letter to Research Councils UK**, undertecknad av åtta framstående professorer i UK. Dessa framhåller bl a inkonsekvensen i att ALPSP å ena sidan hävdar att sällskapen måste tillåtas operera på en fri marknad men å andra sidan vill att RCUK skyddar dem från effekterna av denna marknad.

Stora förlag inför hybridmodeller

Blackwell erbjuder **Online Open**, som startade i februari i år och pågår som försök tom år 2006. Författarna till accepterade artiklar kan välja att betala GBP 1250 och få sina artiklar fritt tillgängliga online via Blackwell Synergy. De inkluderas också i den tryckta upplagan.

Oxford University Press erbjuder **Oxford Open** för 21 tidskifter med start 1 juli i år. **http://www.oxfordjournals.org/jnls/openaccess/about.html**. Man vill skaffa sig erfarenheter och se om konsekvenserna av OA.gynnar större spridning och genomslag av forskningen globalt. OUP har redan en full OA-tidskrift *Nucleic Acids Research*.

Springer Open Choice
Juli 2004 lanserade Springer **Springer Open Choice**. Författarna kan välja att betala USD 3000 för att få sina artiklar fritt tillgängliga via Springer Link.

Jan Velterop blir direktör för Springer's Open Access

Springer har nu rekryterat en av de mest framstående personerna inom OA-rörelsen genom sitt förvärv av Jan Velterop med en lång och välrenommerad erfarenhet av förlagsbranshcen och nu senast

som Publishing Director vid BioMed Central.
Springers VD Derk Haank kommenterar:, “Springer is the first major commercial publisher to provide an Open Access model, making it a pioneer in the industry. We are now taking a further step forward. The appointment of Jan Velterop creates an internal champion for this second component of our publishing policy, making sure Open Access gets the required attention both internally and externally”
Läs mer: **http://www.springeronline.com/sgw/cda/frontpage/0,11855,1-108-2-157191-0,00.html**

OA till offentligfinansierad stamcellsforskning i Kalifornien

University of California Academic Council har enhälligt antagit en Policy on Public Access and Archiving of Research Results som kräver att artiklar som baseras på offentligfinansierad stamcellsforskning blir OA inom sex månader efter accept .

Norge går mot öppna standarder

Den norska regeringen vill att hela den offentliga sektorn skall ha planer klara för övergång till Open Source under 2006. Norges moderniseringsminister, Morton A. Meyer säger att proprietära format inte längre kommer att accepteras i kommunikationen mellan medborgarna och regeringen. Man kommer inte bara att kräva open source utan också öppna standarder. Satsningen skall implementeras under de närmaste tre-fyra åren. **http://europa.eu.int/idabc/en/document/4403/469**

Ny OSI guide till OA publicering för vetenskapliga sällskap

Jan Velterop har skrivit **Guide to Open Access Publishing and Scholarly Societies**, som publicerades av Open Society Institute i juli i år för att ge praktisk hjälp till sällskapen att se över sin publiceringsverksamhet med målet att kunna lägga om till OA, vilket bör passa sällskapens egentliga mission.

Ny SPARC guide för tidskrifter som letar sponsorer

SPARC har publicerat Sponsorships for Nonprofit Scholarly & Scientific Journals: A Guide to Defining & Negotiating Successful Sponsorships. **http://www.arl.org/sparc/resources/pubres.html** - planning Guiden skall vara ett stöd för icke-kommersiella tidskrifter som behöver ekonomiskt bistånd.

Ny OA tidskrift om copyright

Copyright är en ny referentbedömd OA-tidskrift som har tonvikt

på copyright i internetåldern och kommer att täcka aspekter som DRM, internationell copyright, sociala implikationer av copyright etc. I redaktionen ingår auktoriteter som Lawrence Lessig och Michael Geist. Tidskriften kommer att stödja en ny wiki-baserad form av författarsamarbe föutom den mer traditionella. **http://copyrightjournal.org/wiki/index.php/Main_Page**

De ledande forskningsuniversiteten i UK stöder OA

Den s.k. **Russell Group**, bestående av de 19 främsta forskningsintensiva universiteten i UK, har lagt fram ett **Statement on Scholarly Communication and Publishing** som stöder Open Access.

- The Russell Group supports the principle that publicly-funded research should be publicly available.
- The Russell Group believes that the current system of scholarly publishing does not always work in the best interests of the research community.
- Russell Group institutions support the development of institutional repositories of research papers, and will actively encourage their researchers to deposit their work in them.
- The publication charge model of journal publishing is still in its early stages but warrants further consideration. The Russell Group would encourage further research and development on the feasibility of the model.
- The Russell Group supports discussion of these issues in fora such as UUK, RCUK and Government departments, as well as within institutions. The Russell Group consists of 19 major research universities that receive 60% of the research grants in the UK, analogous to the **AAU** in the US.

IFLA stöder A2K fördraget

IFLA gjorde 22 juni ett utmärkt uttalande vid WIPO IIM2 mötet i Geneve **IFLA and the Access to Knowledge (A2K) Treaty**. Utdrag nedan:

- Copyright is not about just protection but was from its early days meant to balance the need to protect creators and entrepreneurs in the work with the user's right to access information and the expression of ideas. The mechanism that makes copyright work is in fact the exceptions and limitations are combined with adequate protection of copyright.
- We call for WIPO to establish global minimum mandatory exceptions and limitations to copyright and related rights because there is an imbalance in power due to the rightsholder having exclusive rights leading to the creation of

monopolies of information. Libraries have a duty to facilitate access to information and knowledge and this does not mean simply making it easy to get permission to use a work for which the user often is required to pay or is otherwise restricted. Libraries also have a duty to support and develop a learning culture, the local and national economy and free civil societies, which means that a certain level of access to information needs to be by right which is what the limitations and exceptions to copyright ensure for the greater public good.

Nyheter och Notiser är sammanställda av **Ingegerd Rabow**, projektledare ScieCom, förste bibliotekarie Lunds Universitets Bibliotek, Biblioteksdirektionen
Ingegerd.Rabow@lub.lu.se